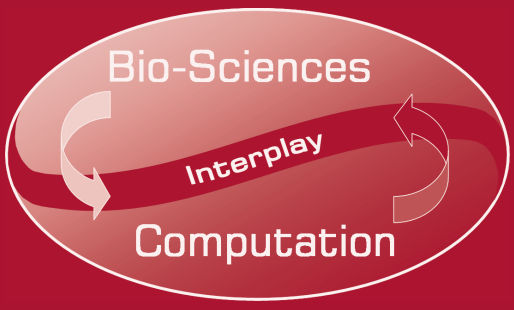
José Manuel Ferrández Vicente
José Ramón Álvarez Sánchez
Félix de la Paz López
Fco. Javier Toledo Moreo (Eds.)

LNCS 7930

Natural and Artificial Models in Computation and Biology

5th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2013 Mallorca, Spain, June 2013, Proceedings, Part I

Part I



Springer

Lecture Notes in Computer Science 7930

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

José Manuel Ferrández Vicente
José Ramón Álvarez Sánchez
Félix de la Paz López
Fco. Javier Toledo Moreo (Eds.)

Natural and Artificial Models in Computation and Biology

5th International Work-Conference on the Interplay
Between Natural and Artificial Computation, IWINAC 2013
Mallorca, Spain, June 10-14, 2013
Proceedings, Part I

Springer

Volume Editors

José Manuel Ferrández Vicente
Fco. Javier Toledo Moreo
Universidad Politécnica de Cartagena
Departamento de Electrónica, Tecnología de Computadoras y Proyectos
Pl. Hospital, 1
30201 Cartagena, Spain
E-mail: {jm.ferrandez; javier.toledo}@upct.es

José Ramón Álvarez Sánchez
Félix de la Paz López
Universidad Nacional de Educación a Distancia
E.T.S. de Ingenería Informática
Departamento de Inteligencia Artificial
Juan del Rosal, 16
28040 Madrid, Spain
E-mail: {jras; delapaz}@dia.uned.es

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-38636-7 e-ISBN 978-3-642-38637-4
DOI 10.1007/978-3-642-38637-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013938949

CR Subject Classification (1998): F.1, F.2, I.2, G.2, I.4-5, J.3-4, J.1

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Living Nature Computing

A trend in computing is the disappearing computer, that is, computers become part of the environment, part of our daily lives, sharing our personal spaces, our amusements, even our own bodies - e.g., embedded systems, accompanying robotics, electronic neuroprosthetics. This implies that computing will also become part of living nature and that computing will be performed by living nature. We believe it is essential to articulate the relationship between different areas of science in order to collaborate and foster a discussion on the interplay of computing with living natural systems.

Natural computing has two sides: computing inspired by nature and computing in nature. These two sides are usually seen as complementary, but our meeting is situated at the intersection/interplay of these two sides. We call this intersection "Living Nature Computing." The term living nature should be taken in a broad sense: from cells via animals to humans. Living nature also puts a focus on natural computing: evolutionary computing and neural networks are examples of natural computing inspired by living nature. DNA computing is included as far as the computation takes place inside living nature. A further focus is the type of computation: we consider engineered computation as inspired by living nature.

Six interesting areas are selected, which share aspects of living nature computing, that we would like to promote in our meetings.

- Cognitive Robotics: Cognitive robotic systems, apart from being practical engineering challenges, are also excellent platforms for experimenting with situated and embodied cognitive theories and models, posing interesting and hard challenges to theoretical and fundamental cognition issues. As Prof. Mira said: "Robotics is the most complete paradigm in Artificial Intelligence: it includes, perception, reasoning, and action."
- Natural Computing: Natural computing refers to computational processes observed in nature, and to human-designed computing inspired by nature. When complex natural phenomena are analyzed in terms of computational processes, our understanding of both nature and the essence of computation is enhanced. Characteristic of human-designed computing inspired by nature is the metaphorical use of concepts, principles, and mechanisms underlying natural systems. Natural computing includes evolutionary algorithms, neural networks, molecular computing, quantum computing, neural modelling, plasticity studies, etc.
- Wetware computation: Wetware computation refers to an organic computer built from living neurons. Silicon-based microchips have physical limits and also power dissipation problems. Wetware computing uses biochemistry

instead of silicon for finding better solutions to be used in future electronics and in information and communications technologies. Molecules or living organisms may carry electrical charge and may perform computing functions with less integration or power dissipation. A wetware computer may be built from leech neurons and be capable of performing simple arithmetic operations and simple pattern-recognition tasks. Another wetware computation is based on cellular cultures. Cells could be neurons from dissociated hippocampus or cortical tissue, neuroblastoma cells, or even PC12 cells, a cell line derived from the rat adrenal medulla.

- Quality of Life Technologies: During the last few years, there is an increasing interest in technologies oriented to improve the quality of life of people. Quality of life technologies (QoLTs) cover a broad area of research including engineering, computer science, medicine, psychology, or social sciences. Typical applications of QoLTs include assistance technologies for people with some kind of disability as, for example, assistive robots, elderly care technologies, or smart homes. However, QoLTs include powerful tools to improve the well-being of individuals and societies in general.
- Biomedical and Industrial Perception Applications: Image understanding is a research area involving both feature extraction and object identification within images from a scene, and a posterior treatment of this information in order to establish relationships between these objects with a specific goal. In biomedical and industrial scenarios, the main purpose of this discipline is, given a visual problem, to manage all aspects of prior knowledge, from study start-up and initiation through data collection, quality control, expert independent interpretation, to the design and development of systems involving image processing capable of tackle with these tasks.
- Web Intelligence and Neuroscience: The study of a user's brainwaves applied to reveal the impact of personalized Web content opens the door for powerful collaborations between Web intelligence and neuroscience. The purpose of this area is to discuss the potential applications of neuroscience methods and concepts in order to evaluate, design, and develop user satisfaction models for intelligent websites. Some of the potential research areas to be analyzed are personalization of Web content and Web application design. On the basis of empirical recordings, it is possible to reveal which features are relevant for each user, therefore learning how to personalize both content and business logic in order to improve the overall user satisfaction. Brainwave analysis can provide the mechanisms to record and measure relevant user behavior patterns regarding memory, attention, and other cognitive aspects. The data obtained by brainwave analysis could guide the development of user satisfaction models, i.e., conceptual and mathematical models that could explain not only why users are satisfied with the provided content, but also predict future satisfaction.

We want to study living nature computing using a systems approach. Computing is not an add-on to the living system, but it is an integral part that is in symbiosis with the system. In fact, it is often not possible to distinguish the

computational parts from the non computational parts. This wider view of the computational paradigm gives us more elbow room to accommodate the results of the interplay between nature and computation. The IWINAC forum thus becomes a methodological approximation (set of intentions, questions, experiments, models, algorithms, mechanisms, explanation procedures, and engineering and computational methods) to the natural and artificial perspectives of the mind embodiment problem, both in humans and in artifacts. This is the philosophy of IWINAC meetings, the "interplay" movement between the natural and the artificial, facing this same problem every two years. This synergistic approach will permit us not only to build new computational systems based on the natural measurable phenomena, but also to understand many of the observable behaviors inherent to natural systems.

The difficulty of building bridges between natural and artificial computation is one of the main motivations for the organization of IWINAC 2013. The IWINAC 2013 proceedings contain the works selected by the Scientific Committee from more than 100 submissions, after the refereeing process. The first volume, entitled *Natural and Artificial Models in Computation and Biology*, includes all the contributions mainly related to the methodological, conceptual, formal, and experimental developments in the fields of neurophysiology and cognitive science. The second volume, entitled *Natural and Artificial Computation in Engineering and Medical Applications*, contains the papers related to bioinspired programming strategies and all the contributions related to the computational solutions to engineering problems in different application domains, especially health applications, including the CYTED "Artificial and Natural Computation for Health" (CANS) research network papers.

An event of the nature of IWINAC 2013 cannot be organized without the collaboration of a group of institutions and people who we would like to thank now, starting with *UNED* and *Universidad Politécnica de Cartagena*. The collaboration of the *UNED Associated Center in Palma de Mallorca* was crucial, as was the efficient work of the Local Organizing Committee, Miguel Angel Vázquez Segura, and Francisco J. Perales López with the close collaboration of the *Universitat de les Illes Balears*. In addition to our universities, we received financial support from the Spanish *CYTED*, *Red Nacional en Computación Naturally Artificial* and *Apliquem Microones 21 s.l.*.

We want to express our gratefulness to our invited speakers, Rodolfo Llinás, from the Department of Physiology and Neuroscience at New York University, Dario Floreano, from the Laboratory of Intelligent Systems at EPFL (Switzerland), and Pedro Gómez-Vilda, from the Oral Communication Lab "Robert Wayne Neucomb" in UPM (Spain), for accepting our invitation and for their magnificent plenary talks.

We would also like to thank the authors for their interest in our call and the effort in preparing the papers, condition sine qua non for these proceedings. We thank the Scientific and Organizing Committees, in particular the members of these committees who acted as effective and efficient referees and as promoters

and managers of pre organized sessions on autonomous and relevant topics under the IWINAC global scope.

Our sincere gratitude goes also to Springer and to Alfred Hofmann and his collaborators, Anna Kramer and Elke Werner, for the continuous receptivity, help efforts, and collaboration in all our joint editorial ventures on the interplay between neuroscience and computation.

Finally, we want to express our special thanks to ESOC, our technical secretariat, and to Victoria Ramos and Nuria Pastor, for making this meeting possible, and for arranging all the details that comprise the organization of this kind of event.

All the authors of papers in this volume, as well as the IWINAC Program and Organizing Committees, would like to commemorate the memory of Professor Mira, who passed away five years ago, both as a great scientist, with an incredible dissemination profile, and as best friend. We still carry his memory deep inside our hearts.

June 2013

José Manuel Ferrández Vicente
José Rámon Álvarez Sánchez
Félix de la Paz López
Fco. Javier Toledo More

Organization

General Chair

José Manuel Ferrández Vicente, UPCT, Spain

Organizing Committee

José Ramón Álvarez Sánchez, UNED, Spain
Félix de la Paz López, UNED, Spain
Fco. Javier Toledo Moreo, UPCT, Spain

Local Organizing Committee

Miguel Ángel Vázquez Segura, UNED, Spain
Francisco J. Perales López, UIB, Spain

Invited Speakers

Rodolfo Llinás, USA
Dario Floreano, Switzerland
Pedro Gómez-Vilda, Spain

Field Editors

Manuel Arias Calleja, Spain
Manuel Luque Gallego, Spain
Juan José Pantrigo, Spain
Antonio Sanz, Spain
Daniel Ruiz, Spain
Javier De Lope Asiaín, Spain
Darío Maravall Gómez-Allende, Spain
Antonio Fernández Caballero, Spain
Rafael Martínez Tomás, Spain
Oscar Martínez Mozos, Spain
María Consuelo Bastida Jumilla, Spain
Rosa María Menchón Lara, Spain
Luis Martínez Otero, Spain
Eduardo Sánchez Vila, Spain

International Scientific Committee

Andy Adamatzky, UK
Michael Affenzeller, Austria
Abraham Ajith, Norway
Amparo Alonso Betanzos, Spain
Jose Ramon Alvarez-Sanchez, Spain
Diego Andina, Spain
Davide Anguita, Italy
Manuel Arias Calleja, Spain
José M. Azorín, Spain
Margarita Bachiller Mayoral, Spain
Antonio Bahamonde, Spain
Dana Ballard, USA
Emilia I. Barakova, The Netherlands
Alvaro Barreiro, Spain
Senen Barro Ameneiro, Spain
M-Consuelo Bastida-Jumilla, Spain
Francisco Bellas, Spain
Guido Bologna, Switzerland
Paula Bonomini, Argentina
Juan Botia, Spain
François Bremond, France
Giorgio Cannata, Italy
Enrique J. Carmona Suarez, Spain
German Castellanos-Dominguez, Colombia
Joaquin Cerda Boluda, Spain
Alexander Cerquera, Colombia
Enric Cervera Mateu, Spain
Antonio Chella, Italy
Santi Chillemi, Italy
Eris Chinellato, Spain
Carlos Colodro-Conde, Spain
Ricardo Contreras, Chile
Erzsebet Csuhaj-Varju, Hungary
Jose Manuel Cuadra Troncoso, Spain
Felix de la Paz Lopez, Spain
Javier de Lope, Spain
Erik De Schutter, Belgium
Angel P. del Pobil, Spain
Ana E. Delgado García, Spain
Jose Dorronsoro, Spain
Gerard Dreyfus, France
Richard Duro, Spain
Reinhard Eckhorn, Germany
Patrizia Fattori, Italy

Juan Pedro Febles Rodriguez, Cuba
Paulo Félix Lamas, Spain
Eduardo Fernandez, Spain
Manuel Fernández Delgado, Spain
Antonio J. Fernández Leiva, Spain
Antonio Fernández-Caballero, Spain
Abel Fernandez-Laborda, Spain
José Manuel Ferrandez, Spain
Kunihiko Fukushima, Japan
Cipriano Galindo, Spain
Cristina Gamallo Solórzano, Spain
Jose A. Gamez, Spain
Jesus Garcia Herrero, Spain
Juan Antonio Garcia Madruga, Spain
Francisco J. Garrigos Guerrero, Spain
Tom D. Gedeon, Australia
Charlotte Gerritsen, The Netherlands
Marian Gheorghe, UK
Pedro Gomez Vilda, Spain
Juan M Gorriz, Spain
Manuel Graña, Spain
Francisco Guil-Reyes, Spain
John Hallam, Denmark
Juan Carlos Herrero, Spain
Cesar Hervas Martinez, Spain
Tom Heskes, The Netherlands
Eduardo Iáñez, Spain
Roberto Iglesias, Spain
Aleksander Igor, UK
Fernando Jimenez Barrionuevo, Spain
Jose M. Juarez, Spain
Joost N. Kok, The Netherlands
Kostadin Koroutchev, Spain
Elka Korutcheva, Spain
Yasuo Kuniyoshi, Japan
Ryo Kurazume, Japan
Petr Lansky, Czech Republic
Jorge Larrey-Ruiz, Spain
Maria Longobardi, Italy
Maria Teresa Lopez Bonal, Spain
Ramon Lopez de Mantaras, Spain
Pablo Lopez Mozas, Spain
Tino Lourens, The Netherlands
Max Lungarella, Japan
Manuel Luque, Spain

Francisco Macia Perez, Spain
George Maistros, UK
Saturnino Maldonado, Spain
Vincenzo Manca, Italy
Daniel Mange, Switzerland
Riccardo Manzotti, Italy
Dario Maravall, Spain
Roque Marin, Spain
Oscar Martinez Mozos, Japan
Rafael Martinez Tomas, Spain
Jose Javier Martinez-Alvarez, Spain
Jesus Medina Moreno, Spain
Rosa-María Menchón-Lara, Spain
Patricio Guisado Miguel Angel, Spain
Jose del R. Millan, Switzerland
Victor Mitrana, Spain
Jose Manuel Molina Lopez, Spain
Javier Monserrat Puchades, Spain
Juan Morales Sanchez, Spain
Federico Moran, Spain
Isabel Navarrete Sanchez, Spain
Taishin Y. Nishida, Japan
Richard A. Normann, USA
Manuel Ojeda-Aciego, Spain
Lucas Paletta, Austria
Jose T. Palma Mendez, Spain
Juan Pantrigo, Spain
Alvaro Pascual-Leone, USA
Gheorghe Paun, Spain
Juan Pazos Sierra, Spain
Francisco Peláez, Brazil
Mario J. Perez Jimenez, Spain
Jose Manuel Perez-Lorenzo, Spain
Franz Pichler, Austria
Maria Pinninghoff, Chile
Jose M. Puerta, Spain
Carlos Puntonet, Spain
Alexis Quesada Arencibia, Spain
Andonie Razvan, USA
Luigi M. Ricciardi, Italy
Mariano Rincon Zamorano, Spain
Victoria Rodellar, Spain
Jose Carlos Rodriguez Rodriguez, Spain
Camino Rodriguez Vela, Spain
Daniel Ruiz, Spain

Ramon Ruiz Merino, Spain
Diego Salas-Gonzalez, Spain
Pedro Salcedo Lagos, Chile
Angel Sanchez, Spain
Eduardo Sánchez Vila, Spain
Juan Vicente Sanchez-Andres, Spain
Jose Santos Reyes, Spain
Antonio Sanz, Spain
Ricardo Sanz, Spain
Shunsuke Sato, Japan
Andreas Schierwagen, Germany
Guido Sciavicco, Spain
Fermin Segovia Roman, Belgium
Radu Serban, The Netherlands
Igor A. Shevelev, Russia
Amari Shun-ichi, Japan
Juan A. Sigüenza, Spain
Jordi Solé i Casals, Spain
Maria Jesus Taboada, Spain
Settimo Termini, Italy
Javier Toledo Moreo, Spain
Rafael Toledo-Moreo, Spain
Jan Treur, The Netherlands
Enric Trillas Ruiz, Spain
Ramiro Varela Arias, Spain
Marley Vellasco, Brazil
Rafael Verdú-Monedero, Spain
Lipo Wang, Singapore
Stefan Wermter, UK
Hujun Yin, UK
Juan Zapata, Spain
Changjiu Zhou, Singapore

Table of Contents – Part I

A Neural Network Simulation of Spreading Depression 1
Daniel Castello Paiva, Diego Andina, and Francisco Javier Ropero Peláez

Robustness of Artificial Metaplasticity Learning to Erroneous Input Distribution Assumptions 9
Marta de Pablos Álvaro and Diego Andina

Neuron Threshold Variability in an Olfactory Model Improves Odorant Discrimination 16
Aaron Montero, Ramon Huerta, and Francisco B. Rodríguez

Contrast Enhancement Mechanisms in the Retinothalamic Circuitry 26
Rubén Ferreiroa and Eduardo Sánchez

Neural Spike Activation in Hippocampal Cultures Using Hebbian Electrical Stimulation 37
V. Lorente, José Manuel Ferrández-Vicente, Fco. Javier Garrigós-Guerrero, Félix de la Paz López, José Manuel Cuadra-Troncoso, José Ramón Álvarez-Sánchez, and Eduardo Fernández

Thalamo-Cortical Network and Seizure Dynamics: A Computational Study 48
F. Vallone, S. Chillemi, M. Barbi, and A. Di Garbo

Modelling Prior and Retrospective Awareness of Actions 62
Dilhan J. Thilakarathne and Jan Treur

Emotional Stress Detection in Contradictory versus Self-consistent Speech by Means of Voice Biometrical Signature 74
Victoria Rodellar-Biarge, Daniel Palacios-Alonso, Elena Bartolomé, and Pedro Gómez-Vilda

Vision Substitution Experiments with See ColOr 83
Guido Bologna, Juan Diego Gomez, and Thierry Pun

Robot Cognitive Stimulation for the Elderly 94
Adriana Tapus and Alexandra-Maria Vieru

The Intelligent Room for Elderly Care 103
Oscar Martinez Mozos, Tokuo Tsuji, Hyunuk Chae, Shunya Kuwahata, YoonSeok Pyo, Tsutomu Hasegawa, Ken'ichi Morooka, and Ryo Kurazume

A Method for Estimating Patient Specific Parameters for Simulation of Tissue Deformation by Finite Element Analysis 113
Ken'ichi Morooka, Shuji Sonoki, Ryo Kurazume, and Tsutomu Hasegawa

REEM Service Robot: How May I Help You? 121
Luca Marchionni, Jordi Pages, Jordi Adell, Jose Rafael Capriles, and Hilario Tomé

Training Study Approaches for a SVM-Based BCI: Adaptation to the Model vs Adaptation to the User 131
Enrique Hortal, Eduardo Iáñez, Andrés Úbeda, José María Azorín, and Eduardo Fernández

Evaluation of a Telepresence Robot for the Elderly: A Spanish Experience 141
Javier Gonzalez-Jimenez, Cipriano Galindo, and Carlos Gutierrez-Castaneda

Empirical Analysis of the Integration of a BCI and an EOG Interface to Control a Robot Arm 151
Eduardo Iáñez, Andrés Úbeda, Enrique Hortal, José María Azorín, and Eduardo Fernández

Interplay between Natural and Artificial Intelligence in Training Autistic Children with Robots 161
Emilia Barakova and Tino Lourens

Use of Multilayer Perceptron vs. Distance Measurement Methods for Classification of Exercises in Telerehabilitation 171
Oscar Marin-Alonso, Daniel Ruiz-Fernández, Antonio Soriano, and Joaquin D. Garcia-Perez

Application of Artificial Metaplasticity Neural Networks to Cardiac Arrhythmias Classification 181
Y. Benchaib, Alexis Marcano-Cedeño, Santiago Torres-Alegre, and Diego Andina

SBRS: Bridging the Gap between Biomedical Research and Clinical Practice 191
Santiago Timón-Reina, Rafael Martínez-Tomás, Mariano Rincón-Zamorano, Tomás García-Sáiz, Estela Díaz-López, and R.M. Molina-Ruíz

Perceptual Analysis of Speech Signals from People with Parkinson's Disease 201
J.R. Orozco-Arroyave, J.D. Arias-Londoño, J.F. Vargas-Bonilla, and Elmar Nöth

Characterization of Speech from Amyotrophic Lateral Sclerosis by Neuromorphic Processing 212
Pedro Gómez-Vilda, Ana Rita M. Londral, José Manuel Ferrández-Vicente, and Victoria Rodellar-Biarge

Fast Approximate Inference in Hybrid Bayesian Networks Using Dynamic Discretisation 225
Helge Langseth, David Marquez, and Martin Neil

An Influence Diagram for the Collaboration in E-learning Environments 235
Antonio R. Anaya and Manuel Luque

A Bayesian Model for Lexical Availability of Chilean High School Students in Mathematics 245
Pedro Salcedo L., Anita Ferreira C., and Fernanda Barrientos C.

Heart Sound Segmentation in Noisy Environments 254
C. Castro Hoyos, S. Murillo-Rendón, and C.G. Castellanos-Domínguez

Kernel Spectral Clustering for Motion Tracking: A First Approach 264
D. Peluffo-Ordóñez, S. García-Vega, and C.G. Castellanos-Domínguez

Multi-labeler Analysis for Bi-class Problems Based on Soft-Margin Support Vector Machines 274
S. Murillo-Rendón, D. Peluffo-Ordóñez, J.D. Arias-Londoño, and C.G. Castellanos-Domínguez

New Cues in Low-Frequency of Speech for Automatic Detection of Parkinson's Disease 283
E.A. Belalcazar-Bolaños, J.R. Orozco-Arroyave, J.F. Vargas-Bonilla, J.D. Arias-Londoño, C.G. Castellanos-Domínguez, and Elmar Nöth

Computer-Aided Detection of Microcalcifications in Digital Mammograms to Support Early Diagnosis of Breast Cancer 293
Nayid Triana and Alexander Cerquera

Influence of Interactions between Virtual Channels in Cochlear Implants 303
Ernesto A. Martínez-Rams and Vicente Garcerán-Hernández

Dealing with Bullying through Genetic Algorithms 313
M. Angélica Pinninghoff, Pedro L. Salcedo, Ricardo Contreras, Andrea Yáñez, and Eduardo Oportus

Using Ant Colony Optimization for Edge Detection in Gray Scale Images 323
Ricardo Contreras, M. Angélica Pinninghoff, and Jaime Ortega

Modeling the Effect of Fixational Eye Movements in Natural Scenes 332
Andrés Olmedo-Payá, Antonio Martínez-Álvarez, Sergio Cuenca-Asensi, José Manuel Ferrández-Vicente, and Eduardo Fernández

A Novel Approach for Quantitative Analysis of 3D Phosphenes 342
C. Soto-Sanchez, Andrés Olmedo-Payá, D. de Santos-Sierra, C. Agullo, and Eduardo Fernández

Inverse Allometry: Foundations for a Bioinspired LVH-Prediction Model 350
María Paula Bonomini, Fernando Ingallina, Valeria Barone, and Pedro Arini

Cochlear Implant: Transcutaneous Transmission Link with OFDM 358
Vicente Garcerán-Hernández and Ernesto A. Martínez-Rams

Deficitary Nervous Excitability and Subjective Contraction of Time: Time-Dispersive Model 368
Isabel Gonzalo-Fonrodona and Miguel A. Porras

Hybrid Tabu Search for Fuzzy Job Shop 376
Juan José Palacios, Jorge Puente, Inés González-Rodríguez, and Camino R. Vela

Statistical Characteristics of Portal Images and Their Influence in Noise Reduction 386
Antonio González-López, María-Consuelo Bastida-Jumilla, Jorge Larrey-Ruiz, and Juan Morales-Sánchez

Self Organized Biogeography Algorithm for Clustering 396
Leila Hamdad, Anissa Achab, Amira Boutouchent, Fodil Dahamni, and Karima Benatchba

Automatic Orientation of Functional Brain Images for Multiplataform Software 406
I. Alvarez Illán, Juan Manuel Górriz, Javier Ramirez, Diego Salas-González, Francisco Jesús Martínez-Murcia, F. Segovia, and C.G. Puntonet

Solving Number Series with Simple Recurrent Networks 412
Stefan Glüge and Andreas Wendemuth

Detection Method for Phase Synchronization in a Population of Spiking Neurons 421
Manuel Lopez and Francisco B. Rodríguez

Author Index 433

Table of Contents – Part II

Ant Colony Algorithms for the Dynamic Vehicle Routing Problem with Time Windows 1
Barry van Veen, Michael Emmerich, Zhiwei Yang, Thomas Bäck, and Joost Kok

Pattern Detection in Images Using LBP-Based Relational Operators 11
José María Molina-Casado and Enrique J. Carmona

Cancer Stem Cell Modeling Using a Cellular Automaton 21
Ángel Monteagudo and José Santos Reyes

Improved Polar Scan-Matching Using an Advanced Line Segmentation Algorithm 32
Israel Navarro Santosjuanes, José Manuel Cuadra-Troncoso, Félix de la Paz López, and Raúl Arnau Prieto

Reactive Navigation and Online SLAM in Autonomous Frontier-Based Exploration 45
Raúl Arnau Prieto, José Manuel Cuadra-Troncoso, José Ramón Álvarez-Sánchez, and Israel Navarro Santosjuanes

Multiscale Dynamic Learning in Cognitive Robotics 56
Pilar Caamaño, Andrés Faíña, Francisco Bellas, and Richard J. Duro

A Vision-Based Dual Anticipatory/Reactive Control Architecture for Indoor Navigation of an Unmanned Aerial Vehicle Using Visual Topological Maps 66
Darío Maravall, Javier de Lope, and Juan Pablo Fuentes Brea

Alignment in Vision-Oriented Syntactic Language Games for Teams of Robots Using Stochastic Regular Grammars and Reinforcement Learning 73
Jack Mario Mingo, Darío Maravall, and Javier de Lope

A Preliminary Auditory Subsystem Based on a Growing Functional Modules Controller 81
Jérôme Leboeuf-Pasquier, Gabriel Fatíh Gómez Ávila, and José Eduardo González Pacheco Oceguera

Robust Multi-sensor System for Mobile Robot Localization 92
A. Canedo-Rodriguez, Víctor Alvarez-Santos, D. Santos-Saavedra, C. Gamallo, M. Fernandez-Delgado, Roberto Iglesias, and C.V. Regueiro

Implicit and Robust Evaluation Methodology for the Evolutionary Design of Feasible Robots 102
Andrés Faíña, Felix Orjales, Francisco Bellas, and Richard J. Duro

Route Learning and Reproduction in a Tour-Guide Robot 112
Víctor Alvarez-Santos, A. Canedo-Rodriguez, Roberto Iglesias, Xosé M. Pardo, and C.V. Regueiro

Meta-ensembles of Classifiers for Alzheimer's Disease Detection Using Independent ROI Features 122
Borja Ayerdi, Alexandre Savio, and Manuel Graña

Results on a Lattice Computing Based Group Analysis of Schizophrenic Patients on Resting State fMRI 131
Darya Chyzhyk and Manuel Graña

Cocaine Dependent Classification on MRI Data Extracting Features from Voxel Based Morphometry 140
M. Termenon, Darya Chyzhyk, Manuel Graña, A. Barros-Loscertales, and C. Avila

A Data Fusion Perspective on Human Motion Analysis Including Multiple Camera Applications 149
Rodrigo Cilla, Miguel A. Patricio, Antonio Berlanga, and José M. Molina

Evaluation of a 3D Video Conference System Based on Multi-camera Motion Parallax 159
Miguel A. Muñoz, Jonatan Martínez, José Pascual Molina, Pascual González, and Antonio Fernández-Caballero

Abandoned Object Detection on Controlled Scenes Using Kinect 169
Antonio Collazos, David Fernández-López, Antonio S. Montemayor, Juan José Pantrigo, and María Luisa Delgado

People Detection in Color and Infrared Video Using HOG and Linear SVM 179
Pablo Tribaldos, Juan Serrano-Cuerda, María T. López, Antonio Fernández-Caballero, and Roberto J. López-Sastre

Smart Spaces and Monitoring Simulation 190
Coral García-Rodríguez, Rafael Martínez-Tomás, and José Manuel Cuadra-Troncoso

Experimental Study of the Stress Level at the Workplace Using an Smart Testbed of Wireless Sensor Networks and Ambient Intelligence Techniques 200
F. Silva, Teresa Olivares, F. Royo, M.A. Vergara, and C. Analide

Towards Usability Evaluation of Multimodal Assistive Technologies Using RGB-D Sensors 210
José Alberto Fuentes, Miguel Oliver, Francisco Montero, Antonio Fernández-Caballero, and Miguel Angel Fernández

Fusion of Overhead and Lateral View Video for Enhanced People Counting 220
Juan Serrano-Cuerda, Marina V. Sokolova, Antonio Fernández-Caballero, María T. López, and José Carlos Castillo

Comparison of Finite Difference and B-Spline Deformable Models in Characterization of 3D Data 230
Rafael Berenguer-Vidal, Rafael Verdú-Monedero, Rosa-María Menchón-Lara, and Álvar Legaz-Aparicio

Automatic Evaluation of Carotid Intima-Media Thickness in Ultrasounds Using Machine Learning 241
Rosa-María Menchón-Lara, María-Consuelo Bastida-Jumilla, Antonio González-López, and José Luis Sancho-Gómez

Active Contours Tool for the Common Carotid Artery Layers Segmentation in Ultrasound Images 250
María-Consuelo Bastida-Jumilla, Rosa-María Menchón-Lara, Juan Morales-Sánchez, and Rafael Berenguer-Vidal

Early Computer Aided Diagnosis of Parkinson's Disease Based on Nearest Neighbor Strategy and *striatum* Activation Threshold 258
Pablo Padilla, Juan Manuel Górriz, Javier Ramírez, Diego Salas-González, and Ignacio Álvarez

Texture Features Based Detection of Parkinson's Disease on DaTSCAN Images 266
Francisco Jesús Martínez-Murcia, Juan Manuel Górriz, Javier Ramírez, I. Alvarez Illán, and C.G. Puntonet

Automatic ROI Selection Using SOM Modelling in Structural Brain MRI 278
Andrés Ortiz García, Juan Manuel Górriz, Javier Ramírez, and Diego Salas-González

Onboard Vision System for Bus Lane Monitoring 286
David Fernández-López, Antonio S. Montemayor, Juan José Pantrigo, María Luisa Delgado, and R. Cabido

Urban Traffic Surveillance in Smart Cities Using Radar Images 296
J. Sánchez-Oro, David Fernández-López, R. Cabido, Antonio S. Montemayor, and Juan José Pantrigo

Vehicle Tracking by Simultaneous Detection and Viewpoint Estimation 306
Ricardo Guerrero-Gómez-Olmedo, Roberto J. López-Sastre, Saturnino Maldonado-Bascón, and Antonio Fernández-Caballero

A Neural Network Approximation of L-MCRS Dynamics for Reinforcement Learning Experiments 317
Jose Manuel Lopéz-Guede, Manuel Graña, Jose Antonio Ramos-Hernanz, and Fernando Oterino

Addressing Remitting Behavior Using an Ordinal Classification Approach 326
Pilar Campoy-Muñoz, P.A. Gutiérrez, and C. Hervás-Martínez

Greedy Sparsification WM Algorithm for Endmember Induction in Hyperspectral Images 336
Ion Marques and Manuel Graña

Dynamic Saliency from Adaptative Whitening 345
Víctor Leborán Alvarez, Antón García-Díaz, Xosé R. Fdez-Vidal, and Xosé M. Pardo

PIR-Based Motion Patterns Classification for AmI Systems 355
Francisco Fernandez-Luque, Juan Zapata, and Ramón Ruiz

Motor Imagery Classification for BCI Using Common Spatial Patterns and Feature Relevance Analysis 365
Luisa F. Velásquez-Martínez, A.M. Álvarez-Meza, and C.G. Castellanos-Domínguez

High-Level Hardware Description of a CNN-Based Algorithm for Short Exposure Stellar Images Processing on a HPRC 375
Jose Javier Martínez-Álvarez, Fco. Javier Garrigós-Guerrero, F. Javier Toledo-Moreo, Carlos Colodro-Conde, Isidro Villó-Pérez, and José Manuel Ferrández-Vicente

Analysis of Connection Schemes between the ICU and the DPU of the NISP Instrument of the Euclid Mission 385
Carlos Colodro-Conde, Rafael Toledo-Moreo, José Javier Díaz-García, Óscar Manuel Tubío-Araujo, Isidro Villó-Pérez, Fco. Javier Toledo Moreo, José Javier Martínez-Álvarez, Fco. Javier Garrigós-Guerrero, José Manuel Ferrández-Vicente, and Rafael Rebolo

Discriminant Splitting of Regions in Traffic Sign Recognition 395
Sergio Lafuente-Arroyo, Roberto J. López-Sastre, Saturnino Maldonado-Bascón, and Rafael Martínez-Tomás

Detection of Fishes in Turbulent Waters Based on Image Analysis 404
Alvaro Rodriguez, Juan R. Rabuñal, Maria Bermudez, and Jeronimo Puertas

Experimental Platform for Accelerate the Training of ANNs with Genetic Algorithm and Embedded System on FPGA 413
Jorge Fe, R.J. Aliaga, and R. Gadea

Segmentation of Weld Regions in Radiographic Images: A Knowledge-Based Efficient Solution 421
Rafael Vilar and Juan Zapata

Automatic Detection of Facial Landmarks in Images with Different Sources of Variations 431
Ángel Sánchez, A. Belén Moreno, and José F. Vélez

An Empirical Study of Actor-Critic Methods for Feedback Controllers of Ball-Screw Drivers 441
Borja Fernandez-Gauna, Igor Ansoategui, Ismael Etxeberria-Agiriano, and Manuel Graña

Robust Solutions for a Robotic Manipulator Optimization Problem 451
Ricardo Soto, Stéphane Caro, Broderick Crawford, and Eric Monfroy

On the Identification and Establishment of Topological Spatial Relations 461
Sergio Miguel-Tomé

Author Index 471

A Neural Network Simulation of Spreading Depression

Daniel Castello Paiva[1], Diego Andina[2], and Francisco Javier Ropero Peláez[3]

[1] University of São Paulo, Brazil
[2] Group for Automation in Signal and Communications, Technical University of Madrid, Spain
[3] Center for Mathematics, Computation and Cognition, UFABC, Brazil
danpaiva@usp.br, andina@gc.ssr.upm.es, francisco.pelaez@ufabc.edu.br

Abstract. With the use of a biologically plausible artificial neural network in which connections are modified through Grossberg's presynaptic learning rule, it is possible to simulate the spreading depression (SD) cortical phenomenon and analyze its behavior depending on different parameters related to neural plasticity and connectivity. The neural network that simulates a simplified cortex is formed by excitatory and inhibitory locally connected neurons. The conditions for the occurrence of SD are analyzed after an external stimulus is applied to the lattice simulating the cortex.

Keywords: Artificial Neural Networks, Computer Simulation, Spreading Depression, Traveling Wave, Neural Plasticity, MatLAB.

1 Introduction

Although the spreading depression (SD) phenomenon was first documented by Leão in 1944[1], the complete understanding of its mechanisms still represents a mystery to neurobiology. SD is a transient wave that travels in all directions on a gray matter substrate after an initial stimulus. The SD phenomenon is very difficult to analyze naturally, due to the fact that it is hard to acquire enough gray matter that can be used in SD experiments. Scientists use, for example, living chicken retina gray matter in order to experiment with SD[11], and each tissue sample can be used only once. Because of this, mathematical models that simulate the phenomenon are very important tools to analyze and study all mechanisms involved in SD without the need of natural tissue.

2 Spreading Depression

Spreading Depression (SD) is a transient wave of suppression of the bioelectrical activity that propagates with constant velocity (2-5 mm/min) [2] [3] in all directions from a stimulated region on a gray matter substrate, like the cortex

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 1–8, 2013.

or hippocampus. Several types of stimuli can be used to experimentally trigger SD [2][4]: chemical stimuli like a drop of a K^+ solution, electrical stimuli as the application of a direct current, or mechanical stimuli like the prick of a needle.

The classical explanation of the SD mechanism[2][3][5][6][7] [8][21] is based on a positive feedback mechanism: the release of potassium ion (K^+) into the interstitium increases with its concentration in such a medium (another theory takes also into account positive feedback involving the excitatory neurotransmitter glutamate released in SD waves). In this way, at least in theory, emotional and physical stress, alcohol or excessive sensory stimulation would lead to intense neuronal excitation of certain regions of the brain thereby causing the accumulation of K^+ in the extracellular space. The membrane potential of neurons in such regions would also increase according to the Nernst equation, thereby depolarizing these cells. Voltage-gated ion channels are opened and K^+ is released from these cells, which become hyperpolarized. Thus, the concentration of K^+ in the interstitial space further increases. By the action of ion channels, these neurons are restored to the polarized state as the time passes. However, in the meantime, K^+ diffuses and affects neighboring neurons, where this process happens again. A propagating wave of K^+ always occurs during SD, regardless of the mechanism of stimulation[8].

Visual disturbances (scotoma) associated with migraine headache, a chronic neurological disorder affecting 10% of the world population[9], can be caused by the propagation of a wave of depressed activity. It is interesting to note that three years before SD was discovered, Lashley[10] argued that a neuronal disturbance (SD) should propagate in order to generate a symptom (scotoma) that propagates as well[5]. Here, we present a computational model of neural network to study this phenomenon.

Some experiments performed in chicken retina, which is an affordable tissue, showed that[3][11][12]:

(i) similarly to an action potential, SD is an all-or-none process;
(ii) there is the annihilation of two waves traveling in opposite directions, after colliding;
(iii) there is wave front recovery after encountering an obstacle; and
(iv) spiral patterns can appear when a wave breaks.

Our model reproduces the first three features. However, our main goal here is to simulate the phenomenon and investigate the influence of the parameter values on the propagation occurrence.

3 Biologically Plausible Artificial Neural Network

The objective of this work is to simulate the spreading depression in a biologically plausible artificial neural network. In order to accomplish this, the concept of complex adaptable responses displayed by natural neurons is modeled with two relevant regulatory or homeostatic mechanisms of neuronal activity that

will be presented: metaplasticity and intrinsic plasticity[13]. For a correct understanding of these two mechanisms, the concept of synaptic plasticity must be first introduced.

3.1 Synaptic Plasticity

Neurotransmitters are molecules responsible for the mediation of the transmission of information between neurons that act on synapses, interacting with ionic channels residing in the synaptic membrane. They allow the inflow or outflow of positive and negative ions such as potassium, sodium, chlorine and calcium. The modulation of the efficacy of information transmission between neurons, being related to the regulation of the number of ionic channels in synapses, is referred to as synaptic plasticity.

Hebb postulated the first model of synaptic plasticity that is known as the Hebb rule [14]. This rule states that when two neurons fire together they wire together. This means that the synaptic strength between neurons with correlated firing tends to increase. Mathematically the change in the synaptic strength (synaptic weight) between neurons i and j is calculated by the product of the output of neuron i and the input I_j (which corresponds to the output of neuron j) multiplied by a learning parameter (Eq. 1).

$$\triangle \omega_{ij} = \xi O_i I_j \tag{1}$$

Some authors proposed revised versions of the Hebb's rule[15], taking into account more recent biological studies. The formulation that was adopted for our model of synaptic plasticity, due to its biological plausibility, is the Grossberg's presynaptic learning rule[16]. The pre-synaptic rule in its incremental version (Eq. 2) is as follows:

$$\triangle \omega_{ij} = \xi I_j (O_i - \omega_{ij}) \tag{2}$$

This incremental version of the pre-synaptic rule is asymptotically equivalent to the following probabilistic version (Eq. 3), where the synaptic weight between two neurons is the conditional probability of the output neuron's firing [17].

$$\omega_{ij} = P(\frac{O_i}{I_j}) \tag{3}$$

3.2 Synaptic Metaplasticity

Synaptic metaplasticity is an important homeostatic mechanism of neurons [18] that slows down the process of weight increment or decrement, making more difficult for the neuron to become either excited or silent. It is also modeled by the presynaptic rule.

Metaplasticity[19] is depicted in Fig.1.a. in which a family of curves represents the variation of weight given the neuron's activation. The value of the synaptic

weight is the parameter that defines what curve must be used. According to this figure, for higher values of the weight the curves are more elongated to the right.

In synapses with high weights, the interval with negative weight variation is broader, thereby favoring synaptic depression. The inverse takes place in curves with lower weights.

Synaptic metaplasticity regulates weight variation because it is a homeostatic mechanism, lowering weight increment in synapses with initial high weights and elevating weight increment in synapses with initial low weights.

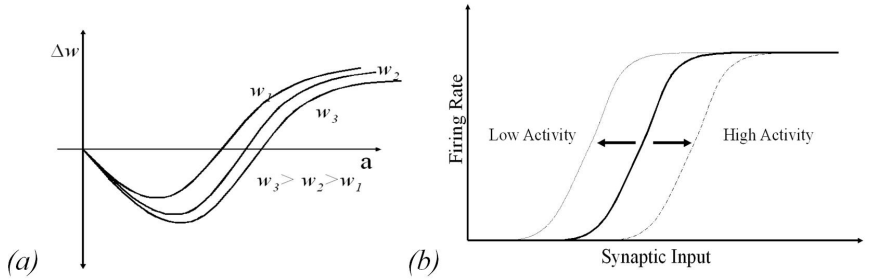


Fig. 1. The curves in (a) represents synaptic metaplasticity, in which more elongated curves are obtained for higher initial synaptic weights; in (b), intrinsic plasticity consists in the shift of the activation function according to higher or lower levels of neural firing

3.3 Intrinsic Plasticity

Although metaplasticity moderates the process of weight variation it does not totally precludes weight annihilation or saturation. To totally avoid either of this two extreme situatoons, another important homeostatic property of real neurons, intrinsic plasticity[20], should be taken into account.

Intrinsic plasticity regulates the position (rightward shift) of the neuron's activation function according to the past average level of neuron's activity. The neuron's activation function is usually modeled as a sigmoidal function, as shown in Eq. 4.

$$O = \frac{1}{1 + e^{-25(a - shift)}} \tag{4}$$

In this equation O represents the output probability of the neuron and a represents the sum of synaptic contributions. The intrinsic plasticity (Fig. 1b) property was modeled according to the following equation that yields the position of the sigmoid in terms of the previous position $shift_{t-1}$ and the previous sigmoid output (Eq. 5).

$$shift_t = \frac{\xi O_{t-1} + shift_{t-1}}{1 + \xi} \tag{5}$$

In this equation the shifting parameter ξ is a small arbitrary factor ranging from 0 to 1. The meaning of this equation is that the more the neuron keeps firing,

the higher will be the rightward shift of the activation function, leading to a moderation of the neuron's firing probability in the future. On the contrary, if the firing probability is low, the sigmoid moves leftwards, increasing the probability of neuron's firing. The shifting parameter resembles a learning factor. If we want quick sigmoid shifting with little interest in accuracy, we select a high value of the shifting parameter. However, if we are mainly interested in accuracy, the shifting parameter must be set to a very small value, for example 0.1 in our model.

4 Computer Model

In order to study the SD, a simplified model of the cerebral cortex was created. This model is composed by a group of 3600 fully inter-connected artificial rate-code neurons, arranged in a lattice of 60 x 60 neurons.

The value of each neuron can vary from total rest (0), to total excitation (1). The value of each neuron depends on its state and the value of its surrounding neurons.

The transition rules are synchronously applied to all neurons on each interaction. However inputs representing an external stimulus can be applied to any neurons in the lattice at any time, triggering the SD.

Each neuron is locally connected to its surrounding neighbors according to a connection matrix of synaptic weights. With this matrix it is possible to connect every neuron from the cortex model with each other using a weight that varies from -1 (full inhibitory connection) to 1 (full excitatory connection). When a connection weight between two neurons is 0, it means that the neurons are not connected.

Every neuron in the model has a certain probability of being inhibitory (we set arbitrarily this number to 30% although other values are also possible) and a 1% probability of being completely inert (not connected with its neighbors). By setting the probability of inhibitory neurons to more than 50%, the SD propagation effect is lost. Between 40% and 50%, there is an incomplete propagation. The ideal propagation happens when the probability of inhibitory neurons ranges between 0% and 30%.

The connection weights are also used to create special parts of the cortex in order to perform several experiments. For example, a physical barrier in which neurons are inert was created in the middle of the cortex by simply setting all connection weights of this region to 0, thereby disconnecting those neurons from all others.

5 Simulations

MatLAB, a commercial software for simulation purposes, was used for the development of the model. The simulations were performed on an Intel I7 high capacity personal computer with a 64 bits operational system. Due to memory limitations (6GB) it was not possible to simulate an experiment using a cortex

with more than 60x60 neurons. Computers with a 64 bit operational system and more than 6 GB of free RAM can easily run the model.

Several simulations were performed in order to represent some SD features. All simulations start with a strong external stimulus applied to neurons placed in a small area of the lattice. This triggers the SD generation that can be observed spreading over the simulated cortex.

Analog to what happens when triggering SD over a gray matter substrate, we use a strong external stimuli to start the process. A small area of the models lattice receives the external stimuli during a certain time interval. Neurons directly affected by stimuli reach high activation values and start triggering their neighbors, which raise their activation values. This process is repeated over time and results in a transient wave that spreads in all directions from the initial stimulus location, as show in figure 2. In figure 2, neurons represented with the darker colors have lower activation values, and neurons represented in lighter colors have higher activation values.

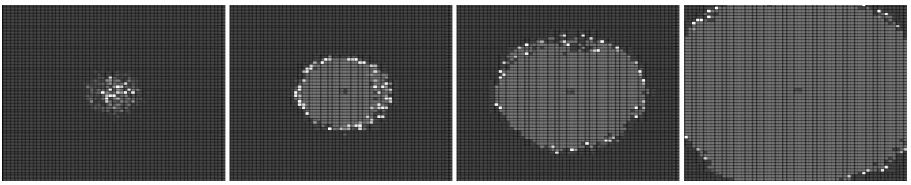


Fig. 2. Visualization of the spreading depression occurring on the simulated cortex

In order to evaluate the model, several configurations were used in the experiments, ranging from special initial conditions to different positions of external stimuli in the lattice. Some of these properties can be compared to those of real experiments, in order to validate the biological plausibility of the SD model.

Some of the del results show similarities with experiments performed in chicken retinas[3][11][12]. One of these properties is the wave annihilation that occurs when two waves traveling in opposite directions collide within the lattice. This behavior of the model is represented in figure 3.

Another behavior simulated in the model is the wave front recovery after the wave encounters an obstacle. In order to represent this, part of the simulated cortex was made inert by preventing the activity of its neurons. The spreading wave, after encountering the barrier, does not stop and simply circumvents the obstacle. This behavior is represented in figure 4.

One important factor to be analyzed within the simulations is the intrinsic plasticity's influence in the overall model behavior. By varying the value of the models shifting parameter, different results were obtained. In all simulations, varying the learning factor in the presynaptic equation does not produce visible alterations in the SD and the generated spreading wave. One factor that influences the speed of the wave propagation is the shifting parameter that represents the speed of the sigmoid shift. By using for this parameter values higher that

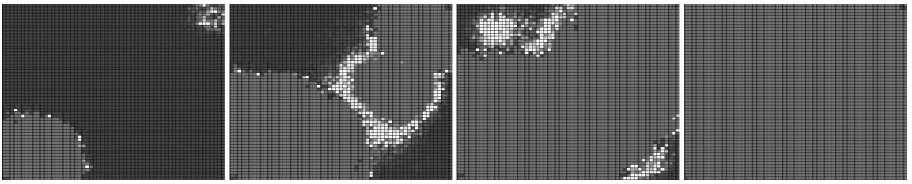


Fig. 3. In the model there is the annihilation of two waves traveling in opposite directions

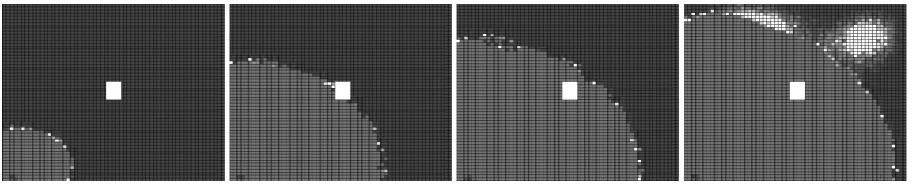


Fig. 4. Visualization of the wave front recovery after encountering an obstacle (white block)

0.4, the wave spreads very quickly along the lattice. If the value is close to 1, all neurons within the lattice activate simultaneously, thus not creating the spreading wave effect. If this parameter value equals to zero, the SD does not occur. So, the optimal value found for the simulations is around 0.1, where constant wave propagation takes place.

6 Conclusions

Some aspects of the SD were simulated using a $60x60$ lattice of biologically plausible neurons. It is possible to observe the SD phenomenon in the simulated cortex after the application of an initial external stimulus, similarly to the SD on a biological cortex.

Analog to the biological occurrence, SD simulations exhibit a constant propagation velocity (with the appropriate selection of parameters).

Different tests were performed to analyze the effect of intrinsic plasticity. They showed that the learning parameter does not influence the wave propagation, but the value of the shifting parameter does influence the speed of the wave. Higher sigmoidal shifting parameters results in higher spreading wave speeds up to the point where all neurons in the lattice discharge at the same time, thus causing the wave effect to disappear.

It is possible to use this model, in addition to other models [21], to investigate possible ways to reduce, or even suppress the occurrences of SD.

References

1. Leão, A.A.P.: Spreading depression of activity in the cerebral cortex. J. Neurophysiol 7, 359–390 (1944)
2. Gorji, A.: Spreading depression: a review of the clinical relevance. Brain Res. Rev. 38, 33–60 (2001)
3. Martins-Ferreira, H., Nedergaard, M., Nicholson, C.: Perspectives on spreading depression. Brain Res. Rev. 32, 215–234 (2000)
4. Weimer, M.S., Hanke, W.: Propagation velocity and triggering threshold of retinal spreading depression are not correlated. Exp. Brain Res. 164, 185–193 (2005)
5. Kager, H., Wadman, W.J., Somjen, G.G.: Simulated seizures and spreading depression in a neuron model incorporating interstitial space and ion concentrations. J. Neurophysiol 512, 495–512 (2000)
6. Dahlem, M.A., Chronicle, E.P.: A computational perspective on migraine aura. Prog. Neurobiol. 74, 351–361 (2004)
7. Somjen, G.G.: Mechanisms of spreading depression and hypoxic spreading depression like depolarization. Physiol. Rev. 81, 1065–1096 (2001)
8. Shapiro, B.E.: Osmotic forces and gap junctions in spreading depression: a computational model. J. Comput. Neurosci. 10, 99–120 (2001)
9. Lipton, R.B., Bigal, M.E.: Migraine: epidemiology, impact, and risk factors for progression. Headache 45(suppl. 1), S3–S13 (2005)
10. Lashley, K.: Patterns of cerebral integration indicated by scotomas of migraine. Arch. Neurol. Psychatry 46, 331–339 (1941)
11. Dahlem, M.A., Muller, S.C.: Self-induced splitting of spiral-shaped spreading depression waves in chicken retina. Exp. Brain Res. 115, 319–324 (1997)
12. Fernandes-de-Lima, V.M., Kogler, J.E., Bennaton, J., Hanke, W.: Wave onset in central gray matter - its intrinsic optical signal and phase transitions in extracellular polymers. An Acad. Bras. Cien. 73, 351–364 (2001)
13. Pelaéz, J.R., Piqueira, J.R.C.: Biological clues for up-to-date artificial neurons. In: Andina, D., Phan, D.T. (eds.) Computational Intelligence: for Engineering and Manufacturing, vol. 1, pp. 1–19. Springer, Berlin (2006)
14. Hebb, D.O.: The organization of behavior. Wiley, New York (1949)
15. Brown, T.H., Kairiss, E.W., Keenan, C.L.: Hebbian Synapses: Biophysical Mechanisms and Algorithms. Annual Review in Neurosciences 13, 475–511 (1990)
16. Grossberg, S.: Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics 23, 121–134 (1976)
17. Minai, A.A., Levy, W.B.: The Dynamics of Sparse Random Networks. Biological Cybernetics 70(2), 177–187 (1993)
18. Desai, N.: Homeostatic plasticity in the CNS: synaptic and intrinsic forms. Journal of Physiology 97, 391–402 (2003)
19. Pelaéz, J.R., Simóes, M.G.: A Computational Model of Synaptic Metaplasticity. In: Proceedings of the International Joint Conference of Neural Networks, vol. 1, pp. 6–11 (1999)
20. Desai, N.S., Rutherford, L.C., Turrigiano, G.: Plasticity in the intrinsic excitability of cortical pyramidal neurons. Nature Neuroscience 2(6), 515–520 (1999)
21. Monteiro, L.H.A., Paiva, D.C., Piqueira, J.R.C.: Spreading depression in mainly locally connected cellular automaton. Journal of Biological Systems 14(4), 617–629 (2006)

Robustness of Artificial Metaplasticity Learning to Erroneous Input Distribution Assumptions

Marta de Pablos Álvaro and Diego Andina

Group for Automation in Signals and Communications,
Technical University of Madrid

Abstract. Artificial Metaplasticity learning algorithm is inspired by the biological metaplasticity property of neurons and Shannon's information theory. In this research, Artificial Metaplasticity on multilayer perceptron (AMMLP) is compared with regular Backpropagation by using input sets generated with different probability distributions: Gaussian, Exponential, Uniform and Rayleigh. Artificial Metaplasticity shows better results than regular Backpropagation for Gaussian and Uniform distribution while regular Backpropagation shows better results for Exponential and Rayleigh distributions.

1 Introduction

Synaptic plasticity is the ability of the synapse between two neurons to modulate its efficiency. This plasticity involves different cellular processes that modify the synaptic function. These changes in synaptic efficiency can cause an enhancement in synaptic strength (Long-term Potentiation, LTP) or a reduction in synaptic strength (Long-term Depression, LTD) [1, 2].

W.C.Abraham [3] introduced the concept of biological metaplasticity as a higher level of plasticity and therefore, as a change in the way synaptic efficacy is modified. Metaplasticity is a biological concept related to memory and learning processes [4] and widely used in the fields of biology, neuroscience, physiology, neurology and others [5–7].

Artificial Metaplasticity (AMP) term was first introduced by Andina et al [8] for an Artificial Neural Network (ANN)of the Multilayer Perceptron type (MLP), referred as AMMLP. The implementation and application of AMMLP trained by Backpropagation Algorithm (BP) was presented in [8] and some of the successful applications are detailed in [9–11].

Regarding all AMP models implemented so far, the model that matches Metaplasticity with Sannon information theory is the most efficient in terms of learning and performance. This model establishes that less frequent patterns give more information than the most frequent ones.

Hence, during the training phase, the AMMLP assigns higher values for updating the weights in the less probable activations than in the ones with higher probability and therefore, the weights are updated according to the probability of the input patterns.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 9–15, 2013.

In Artificial Metaplasticity a weighting function f(x), related to the probability distribution of the input vectors is applied to the error function to be minimized. In order to model the probability density function of the input patterns, a suboptimal function is used because the optimal function is not known *a priori*. Therefore, the training pattern distribution is approximated as a Gaussian distribution.

$$f_X^*(x) = \frac{A}{\sqrt{(2\pi)^N}.e^{B\sum_{i=1}^{N} x_i^2}} \tag{1}$$

being N the number of components of the input vector X that feeds the first hidden layer. For the second hidden layer, X is substituted for the output vector of the first hidden layer and so on. Parameters A and B are empirically obtained.

2 Multilayer Perceptron Structure

In order to compare Artificial Metaplasticity with regular Backpropagation, a multilayer perceptron with the following features has been simulated:

- Number of input neurons: 2 (number of attributes of the database)
- Number of hidden layers: 1
- Number of hidden neurons: 2
- Number of output neurons: 1
- Activation function: sigmoid function with values between [0,1]
- Initial weights: random values in [-0.6, 0.6] interval

The terminology for each weight of the network is wL(i,j), being:

- L: layer, L=1 for hidden layer and L=2 for output layer
- i: input, (1,2) for L=1 and (1,2) for L=2
- j: output, (1,2) for L=1 and (1) for L=2
- dwL(i,j): variation of wL(i,j) weight obtained in one training iteration

The terminology for each neuron of the network is nL(i) being:

- L: layer, L=1 for hidden layer and L=2 for output layer
- i: neuron, (1,2) for L=1 and (1) for L=2

A scheme of this neural network is shown in figure below:

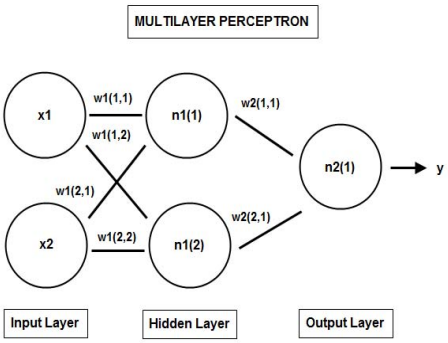


Fig. 1. Multilayer Perceptron used in this research

3 Results

This section presents the results obtained by comparing AMMLP with Backpropagation.

First of all, parameters of both methods were optimized for this multilayer perceptron structure, finding the following values:

- Artificial Metaplasticity: weighting function parameters: $A = 0.5$ and $B = 0.35$
- Backpropagation: learning ratio $\eta = 50$

These parameters were empirically optimized by repeating the experiments with different values and selecting the values that generated best results.

The aim of the experiments is to compare the results obtained with Artificial Metaplasticity and regular Backpropagation by using input sets generated with different probability distributions. The probability distributions used for the experiments were: Gaussian, Exponential, Uniform and Rayleigh.

The experiments consist on training the network until the error is below 0.3, at this point the network has started to learn but there is still margin for the end of the learning. Each experiment is repeated 50 times in order to obtain the average number of iterations needed to achieve an error below 0.3.

3.1 Gaussian Distribution

A two-dimensional input set of eight elements is generated with Matlab using a Gaussian distribution with mean 0 and variance 1. The output set is calculated

as the result of evaluating if the sum of the two dimensions of each element is above 1. Comparison of Backpropagation and Artificial Metaplasticity is shown in the table below:

Table 1. Gaussian Distribution

	Backpropagation					Artificial Metaplasticity			
Errors	0.3750	0.3750	0.3750	0.3750	0.1250	0.6250	0.6250	0.4375	0.2500
dw1(1,1)	-0.0023	-0.0010	-0.0049	-0.0361	-0.0177	-0.0309	-0.0286	0.0451	0.0223
dw1(1,2)	0.0009	0.0004	0.0039	0.0509	0.0089	-0.3470	-0.2534	-0.0753	-0.0204
dw1(2,1)	0.0028	0.0011	0.0059	0.0431	0.0212	-0.0665	-0.0616	0.0971	0.0480
dw1(2,2)	-0.0011	-0.0005	-0.0047	-0.0609	-0.0107	-0.7468	-0.5454	-0.1621	-0.0439
dw2(1,1)	-0.0081	-0.0017	-0.0064	-0.0130	-0.0083	2.0895	1.4078	0.4477	0.1208
dw2(2,1)	-0.0032	-0.0008	-0.0052	-0.0931	-0.0036	1.4972	0.4919	0.0720	0.0114
w1	-3.9933 -3.1234 5.4884 3.1668					0.2493 0.0409 -3.5997 -1.8993			
w2	-5.7263 4.5436					0.5560 -3.9251			
n1(1)	0.6139	0.5070	0.3896	0.1080	0.3934	0.4571	0.3093	0.2096	0.1975
n1(2)	0.2428	0.2313	0.3128	0.7715	0.1731	0.3275	0.1081	0.0337	0.0186
n2(1)	0.1225	0.0595	0.1384	0.8211	0.1579	0.3003	0.3763	0.6828	0.8477
Iterations	2,7					1,7			

3.2 Exponential Distribution

A two-dimensional input set of eight elements is generated with Matlab using an Exponential distribution with parameter with value 2 for one of the dimensions and value 4 for the other dimension. The output set is calculated as the result of evaluating if the sum of the two dimensions of each element is above 0.4.

Table 2. Exponential Distribution

	Backpropagation	Artificial Metaplasticity
Errors	0.2500	0.5000 0.3750 0.1875
dw1(1,1)	-0.0003539	-0.0705 -0.0108 -0.0114
dw1(1,2)	-0.0001802	0.0058 0.0043 0.0040
dw1(2,1)	-0.0003539	-0.0705 -0.0108 -0.0114
dw1(2,2)	-0.0001802	0.0058 0.0043 0.0040
dw2(1,1)	-0.0012	-0.1566 -0.0742 -0.1167
dw2(2,1)	-0.0004411	-0.0194 -0.0025 -0.0014
w1	0.2505 0.4552 1.0428 0.9998	-2.4331 -1.8222 2.0201 2.0201
w2	-0.3653 0.4208	-3.3080 2.9226
n1(1)	0.1523	0.7494 0.9457 0.9701
n1(2)	0.0538	0.0926 0.0325 0.0116
n2(1)	0.0952	0.0798 0.0481 0.0599
Iterations	1,1	3,4

3.3 Rayleigh Distribution

A two-dimensional input set of eight elements is generated with Matlab using a Rayleigh distribution with parameter with value 1. The output set is calculated as the result of evaluating if the sum of the two dimensions of each element is above 2.

Table 3. Rayleigh Distribution

	Backpropagation	Artificial Metaplasticity
Errors	0.3750 0.3750 0.2500	0.5000 0.5000 0.3750 0.1250
dw1(1,1)	-0.0313 -0.0001 -0.0000	-0.1033 0.0571 0.1364 0.0438
dw1(1,2)	-0.0277 -0.0002 -0.0001	0.1399 0.3776 0.7408 0.1766
dw1(2,1)	0.0103 0.0000 0.0000	0.3061 -0.1692 -0.4040 -0.1297
dw1(2,2)	0.0091 0.0001 0.0000	-0.4146 -1.1188 -2.1947 -0.5232
dw2(1,1)	0.0339 0.0001 0.0000	-2.5367 -1.7799 -1.4872 -0.1940
dw2(2,1)	0.0343 0.0002 0.0000	-1.4857 -1.7464 -2.2365 -0.4088
w1	-2.5113 -2.4465 -2.1449 -2.3469	-0.5802 -1.3558 -4.0573 -3.8417
w2	-2.6838 -2.3420	-1.1386 -4.6187
n1(1)	0.2629 0.0316 0.0158	0.5716 0.4065 0.3338 0.3312
n1(2)	0.2656 0.0481 0.0266	0.3348 0.3989 0.5020 0.6979
n2(1)	0.2030 0.9365 0.9693	0.6906 0.6147 0.6718 0.1515
Iterations	2,6	3,8

3.4 Uniform Distribution

A two-dimensional input set of eight elements is generated with Matlab using a Uniform distribution in [0, 1] interval. The output set is calculated as the result of evaluating if the sum of the two dimensions of each element is above 1.

Table 4. Uniform Distribution

	Backpropagation	Artificial Metaplasticity
Errors	0.3750 0.3750 0.2500	0.3750 0.3750 0.1250
dw1(1,1)	-0.0073 -0.0072 -0.0088	-0.2728 -0.2965 -0.5624
dw1(1,2)	-0.0038 -0.0014 0.0044	-0.0387 0.1224 1.1263
dw1(2,1)	-0.0048 -0.0047 -0.0058	0.3945 0.4287 0.8132
dw1(2,2)	-0.0025 -0.0009 0.0029	0.0559 -0.1770 -1.6285
dw2(1,1)	0.0087 0.0034 0.0019	-0.7304 -0.2967 -0.3817
dw2(2,1)	0.0109 0.0062 0.0225	-0.8786 -1.0822 -2.8398
w1	-2.7857 -2.3737 0.0975 0.6735	0.3065 2.8794 -0.7570 -3.5916
w2	-4.6405 0.7925	2.4519 -4.1179
n1(1)	0.1560 0.0822 0.0132	0.4527 0.1636 0.1026
n1(2)	0.1946 0.1503 0.1563	0.5445 0.5968 0.7634
n2(1)	0.0638 0.0451 0.3980	0.9412 0.9327 0.4811
Iterations	3,6	2,6

3.5 Comparison

In the following table are summarized the results obtained for each distribution:

Table 5. AMMLP and BP Comparison

Distribution	Backpropagation	Artificial Metaplasticity
Gaussian	2,7	1,7
Exponential	1,1	3,4
Rayleigh	2,6	3,8
Uniform	3,6	2,6

Artificial Metaplasticity shows better results than Backpropagation for Gaussian and Uniform distribution but Backpropagation method shows better results for Exponential and Rayleigh distribution. Backpropagation shows the best result with Exponential distribution while Artificial Metaplasticity shows the best result with Gaussian distribution.

4 Discussion

The results show that AMMLP algorithm using a Gaussian weighting function achieves better results than regular Backpropagation when the input follows a Gaussian distribution. Gaussian distribution occurs very frequently in statistics, economics, natural and social sciences so this result has big implications and multiple applications. This result agrees with [9–11] where AMMLP is tested on different multidisciplinary applications showing a much more efficient training than the best systems previously used in the state of the art.

For the input with uniform distribution the results obtained with AMMLP are not so good as for the Gaussian distribution input but are still better than results obtained with regular Backpropagation. However, regular Backpropagation shows better results than AMMLP when the input follows a Rayleigh or an Exponential distribution. A proposal for these distributions is to use the output of the neural network as the AMMLP weighting function instead of the Gaussian weighting function and test the results.

From tables 1 to 4 it is observed that the weight variations of both weight matrix (w1 and w2) are much higher when AMMLP algorithm is used. In regular Backpropagation the initial random weight is slowly modified through low weight variations while AMMLP initial random weight is modified faster with higher weight variations.

This research has been developed using a simple multilayer perceptron structure so it would be convenient to use a more complex structure with a higher number of neurons in order to test results.

5 Conclusions

In this paper, artificial metaplasticity on multilayer perceptron (AMMLP) is compared with regular Backpropagation using different probability distribution inputs. It is concluded that AMMLP obtains better results than regular Backpropagation for Gaussian and Uniform distribution inputs. This result has big implications and multiple applications due to the high frequency of Gaussian distribution appearance in statistics, economics, natural and social sciences.

References

1. Jedlicka, P.: Synaptic plasticity, metaplasticity and BCM theory 103(4), 137–143 (2002)
2. Malenka, R., Bear, M.: LTP and LTD: An Embarrassment of Riches. Neuron 44(1), 5–21 (2004), doi:10.1016/j.neuron.2004.09.012
3. Abraham, W.C.: Activity-dependent regulation of synaptic plasticity (metaplasticity) in the hippocampus. In: Kato, N. (ed.) The Hippocampus: Functions and Clinical Relevance, pp. 15–26. Elsevier, Amsterdam (1996)
4. Abraham, W.C.: Metaplasticity: Key Element in Memory and Learning? News in Physiological Sciences 14(2), 85 (1999)
5. Abraham, W.C., Bear, M.F.: Metaplasticity: The plasticity of synaptic plasticity. Trends in Neurosciences 19, 126–130 (1996), doi:10.1016/S0166-2236(96)80018-X
6. Kinto, E.A., Del Moral Hernandez, E., Marcano, A., Ropero Peláez, J.: A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 628–636. Springer, Heidelberg (2007)
7. Chiappalone, M., Vato, A., Berdondini, L., Koudelka, M., Martinoia, S.: Network Dynamics and Synchronous Activity in Cultured Cortical Neurons. Int. J. Neural Syst. 17(2), 87–103 (2007)
8. Andina, D., Alvarez-Vellisco, A., Jevtic, A., Fombellida, J.: Artificial metaplasticity can improve artificial neural network learning. Intelligent Automation and Soft Computing, SI on Signal Processing and Soft Computing 15, 683–696 (2009) ISSN: 1079-8587
9. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Andina, D.: Breast cancer classification applying artificial metaplasticity algorithm. Neurocomputing, doi:10.1016/j.neucom.2010.07.019
10. Marcano-Cedeño, A., Martín de la Bárcena, A., Jiménez-Trillo, J., Piñuela, J.A., Andina, D.: Artificial Metaplasticity Neural Network Applied to Credit Scoring. Int. J. Neural Syst. 21(4), 311–317 (2011), doi:10.1142/S0129065711002857
11. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Andina, D.: Wood Defects Classification Using Artificial Metaplasticity Neural Network. In: Proc. 35th Annual Conf. on of the IEEE Industrial Electronics Society, Porto, Portugal, pp. 3422–3427 (2009), doi:10.1109/IECON.2009.5415189

Neuron Threshold Variability in an Olfactory Model Improves Odorant Discrimination

Aaron Montero[1], Ramon Huerta[1,2], and Francisco B. Rodríguez[1]

[1] Grupo de Neurocomputación Biológica, Dpto. de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[2] BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093-0402, USA
aaron.montero@uam.es

Abstract. We used a model based on the olfactory system of insects to analyze the impact of neuron threshold variability in the mushroom body (MB) for odorant discrimination purposes. This model is a single-hidden-layer neural network (SLN) where the input layer represents the antennal lobe (AL), which contains a binary code for each odorant; the hidden layer that represents the Kenyon cells (KC) and the output layer named the output neurons. The KC and output layers are responsible for learning odor discrimination. The binary code obtained for each odorant in the output layer has been used to measure the discrimination error and to know what kind of thresholds (heterogeneous or homogeneous) provide better results when they are used in KC and output neurons. We show that discrimination error is lower for heterogeneous thresholds than for homogeneous thresholds.

Keywords: neural variability, pattern recognition, odor learning and discrimination, decision making.

1 Introduction

Neural thresholds vary in olfactory receptor neurons (ORN) [1] and in Kenyon cells (KCS) in the insect mushroom body (MB) [14]. Similarly electronic noses [3] also utilize different detection thresholds for different odorants to improve gas discrimination. Neural variability in the form of threshold variability is a general property of neurons in the brain. Is the threshold variability an evolutionary advantage which allows better odorant discrimination, or does it happen because there is a biological impossibility to produce the same threshold for each neuron?

To investigate if neural threshold variability increases odorant discrimination performance, we use a simple model of the olfactory system of insects [7,8] based on McCulloch-Pitts neurons [13]. The insect olfactory pathway starts at the antenna, where a massive number of receptors encode the odor stimulus in a high-dimensional code. In locusts [10], this number is approximately 100,000

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 16–25, 2013.

neurons. This information is then sent to the AL for additional processing. In the locust, the AL approximately has 1,000 neurons that compress the ORN information. The AL exhibits complex dynamics produced by the interaction of its excitatory and inhibitory neural populations [2,5,11]. The excitatory cells are called projection neurons because they only transmit the result of AL computation to deeper regions. The projection neurons deliver the AL output to the 50,000 cells of the MB, KC, in a fan-out phase which increases the separability between different odor encodings. This fan-out phase combined with the sparse firing for these KC [14,16,6] facilitate the odorant discrimination process, which is realized by output neurons in a fan-in phase.

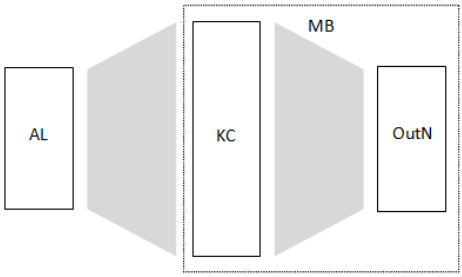


Fig. 1. The structure of the model is divided into two parts: antennal lobe (AL) and mushroom body (MB). MB is divided into two additional layers: Kenyon cells (KC) and output neurons (OutN).

We focus on the AL and MB (Fig.1), where the input to SLN is AL, which is connected to MB through a non-specific connectivity matrix. The reason for this non-specific connectivity matrix is due the individual connection variability of insects of the same species [12,15]. The other layers of the SLN, hidden and output, are composed by KC and output neurons respectively, which are connected by a connectivity matrix that implements Hebbian-like learning [4].

Our goal is to compare homogeneous and heterogeneous thresholds to determine whether neural variability improves odorant discrimination. To investigate that, we measure the discrimination error obtained in the output layer. This discrimination error represents the percentage of odorants which have not been correctly classified. A correctly classified odorant always generates the same output pattern A' for a certain input pattern A. Furthermore, we calculate the percentage of KC spikes to prove that discrimination success is related to the sparseness condition observed in the KC layer.

We conclude that odorant discrimination improves with neuron threshold variability and that the discrimination performance is closely related to sparse activity of the KC population.

2 Olfactory Model and Methods

2.1 Neuron Model

In locusts, activity patterns in AL are practically time-discretized by a periodic feedforward inhibition onto MB calyxes, and activity levels in KCs are very low [14]. Thus, information is represented by time-discrete, sparse activity patterns in MB in which each KC fires at most once in each 50 ms local field potential oscillation cycle. Because of this intermittent discrete activity, we have used the McCulloch-Pitts model [13] in all neurons of the hidden and output layers. This neuron model uses the threshold step function as activation function. Therefore, we have the following:

$$y_j = \Theta(\sum_{i=1}^{N_{AL}} c_{ji}x_i - \theta_j), \qquad z_l = \Theta(\sum_{j=1}^{N_{KC}} w_{lj}y_j - \varepsilon_l) \tag{1}$$

where x_i, y_j and z_l are activation states for a input, hidden and output neuron respectively, c_{ji} and w_{lj} are weights which links two neurons, θ_j and ε_l are thresholds for the hidden and output neuron respectively, and Θ is the activation function.

2.2 Network Model

The network model is a SLN with an input layer with 100 neurons, a hidden layer with 5000 neurons (locust has a ratio of 1:50 between neurons of the input and hidden layer) and an output layer with 10 neurons [8]. These dimensions were chosen because they ensure a high probability of discrimination for the input used [7] for a relatively low computational cost.

The connectivity matrices, C and W, are initialized at the beginning of each odorant discrimination process. We generate a matrix with random values uniformly distributed, $[0, 1]$, with the same dimensions as our connectivity matrix. We establish connections in our connectivity matrix using the probability of connection matrix, p_c and p_w, as a threshold on the values of the random matrix: If a float value is equal or less than the probability of connection, one connection is established, otherwise no connection. In the case of the connectivity matrix C, this configuration remains fixed throughout the odorant discrimination process. However, for the connectivity matrix W, its configuration will be updated using Hebbian learning.

Finally, we have to mention that the synaptic model of this network is completely binary. Therefore, activation states for a neuron and weights can only take values of 0 or 1.

2.3 Hebbian Learning

As mentioned above, the connectivity matrix W, which links KC and output neurons, has olfactory associative learning, which can be simulated by using

Hebbian learning [4]. It allows the strengthening or weakening the connections of a connectivity matrix, as follows [7,8]:

$$
\begin{aligned}
& w_{lj}(t+1) = H(z_l, y_j, w_{lj}(t)), \\
& H(1,1,w_{lj}(t)) = \begin{cases} 1 & \textit{with probability } p_+, \\ w_{lj}(t) & \textit{with probability } 1-p_+, \end{cases} \\
& H(1,0,w_{lj}(t)) = \begin{cases} 0 & \textit{with probability } p_-, \\ w_{lj}(t) & \textit{with probability } 1-p_-, \end{cases} \\
& H(0,1,w_{lj}(t)) = w_{lj}(t), \qquad H(0,0,w_{lj}(t)) = w_{lj}(t).
\end{aligned}
\tag{2}
$$

where the future connection state $w_{lj}(n+1)$ is determined by a function $H(z_l, y_j, w_{lj}(t))$, which depends on the output layer neuron z_l, the hidden layer neuron y_j and the current connection state $w_{lj}(n)$. If the output layer neuron has not fired, the connection state is not changed. However, if the output layer neuron has fired, the connection state depends on the hidden layer in the following ways:

- If the hidden layer neuron has fired, then the connection between these neurons is created with a probability p_+.
- If the hidden layer neuron has not fired, then the connection between these neurons is destroyed with a probability p_-.

2.4 Odorants

The odorants used in our model as input have no correspondence to real odorant data. Instead, we have used orthogonal and character-based encodings (Fig.2). The reason for using these encodings is because we need these odorants have a minimum distance between them to ensure they are different odorants. In the case of orthogonal encoding, we have wanted to observe what happens if different odorants do not share activated neurons in the input layer, AL. We have used orthogonal encodings all of which have the same number of active neurons in the input layer. Also, this number of active neurons is maximal. Therefore, if we have 100 input neurons, we will have 10 active neurons for 10 different odorant patterns, and 20 for 5 different odorant patterns. In the case of character-based encodings, we have used numerical characters represented in matrices of dimensions 10×10, which we have later converted to vectors of dimension 100. The minimum Hamming distance between these encodings is 4 activated neurons, for the numerical characters 5 and 6.

We have worked with four set of odorants, they have been created from a initial set of odorants, which have been replicated three times and introduced them some noise. This noise represents a set of input neurons which have changed their state of activation (active/inactive). We have used a noise that affects a

specific number of neurons, which is in proportion (20%) to the number of active neurons. The four set of odorants are as follows:

- 15 Orthogonal odorants: 5 orthogonal odorant patterns (20 active neurons) repeated 3 times with noise in 4 neurons.
- 15 Character-based odorants: 5 character-based odorant patterns (with a minimum of 28 active neurons) repeated 3 times with noise in 6 neurons.
- 30 Orthogonal odorants: 10 orthogonal odorant patterns (10 active neurons) repeated 3 times with noise in 2 neurons.
- 30 Character-based odorants: 10 character-based odorant patterns (with a minimum of 28 active neurons) repeated 3 times with noise in 6 neurons.

We have used these odorants with noise to observe if they can be well classified despite noise, after we have known the discrimination error when these odorants are presented without noise. This error for odorant discrimination in the absence of noise will be shown in the results section.

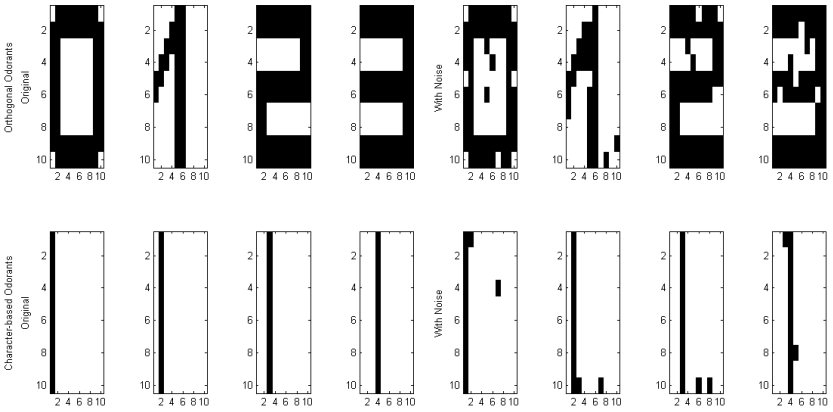


Fig. 2. Examples of orthogonal (top panels) and character-based (bottom panels) encodings without and with noise. Colours: black (1, active neuron), white (0, inactive neuron).

2.5 Limit Thresholds

A limit threshold is the minimum threshold value which prevents a neuron from spiking for an odorant. This value has been used as threshold in our neurons, in order to prove how important threshold variability is in the problem of odorant discrimination. This limit threshold is calculated for each neuron and each odorant as follows:

$$\theta_j^O = \sum_{i=1}^{N_{AL}} c_{ji} x_i^O, \qquad \varepsilon_l^O = \sum_{j=1}^{N_{KC}} c_{lj} y_j^O \tag{3}$$

where neuron j spikes $\forall\theta_j, 0 \leq \theta_j^O < \theta_j$, and neuron l spikes $\forall\varepsilon_l, 0 \leq \varepsilon_l^O < \varepsilon_l$. Being θ_j^O the limit threshold for a KC (j) and an odorant (O), and ε_l^O the limit threshold for an output neuron (l) and an odorant (O), and both are natural numbers. These thresholds are calculated only one time in the odorant discrimination process, both before Hebbian learning. Therefore, the limit threshold matrix, which stores all limit threshold of a layer, for the hidden layer has dimension $N_{KC} \times N_{ODOR}$, and dimension $N_{OutN} \times N_{ODOR}$ for the output layer.

$$\theta_j^O = \begin{pmatrix} \theta_1^1 & \cdots & \theta_1^{N_{ODOR}} \\ \vdots & \ddots & \vdots \\ \theta_{N_{KC}}^1 & \cdots & \theta_{N_{KC}}^{N_{ODOR}} \end{pmatrix} \qquad \varepsilon_l^O = \begin{pmatrix} \varepsilon_1^1 & \cdots & \varepsilon_1^{N_{ODOR}} \\ \vdots & \ddots & \vdots \\ \varepsilon_{N_{OutN}}^1 & \cdots & \varepsilon_{N_{OutN}}^{N_{ODOR}} \end{pmatrix} \tag{4}$$

The purpose of these matrices (Fig.3) is to know all possible thresholds for each layer and choose the one which improves odorant discrimination.

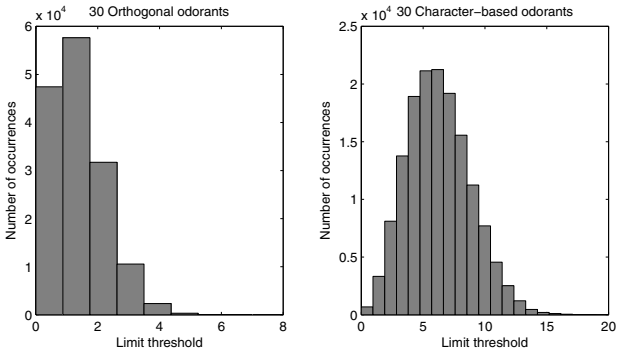


Fig. 3. Example of limit threshold distributions in KC for different odorant sets: 30 Orthogonal odorants (left panel), 30 Number odorants (right panel)

As we have already explained the meaning of limit threshold, we will explain how they are calculated for the cases of homogeneous and heterogeneous thresholds.

Homogeneous Thresholds. To calculate the homogeneous thresholds, we obtain the limit threshold matrix for the hidden layer and we take the minimum and maximum of this matrix. We take all values between the minimum and maximum, including these, to use as thresholds for the hidden layer. The aim of this process is to obtain the minimum discrimination error for each threshold and the spike rate for this minimum.

To achieve this minimum discrimination error, for each hidden layer threshold we obtain its maximum inhibitory matrix for the output layer and we take the minimum and maximum of this matrix. We calculate the discrimination error for all possible combinations and take the minimum observed. This value is the minimum discrimination error for a hidden layer threshold.

Heterogeneous Thresholds. In the case of heterogeneous thresholds, we obtain the limit threshold matrix for the hidden layer but we do not take all possible combinations. We obtain the distribution of limit thresholds for each hidden layer neuron and we select the value which prevents each neuron from firing for a certain percentage of odorants. These values will be the limit thresholds for these neurons (Fig.4).

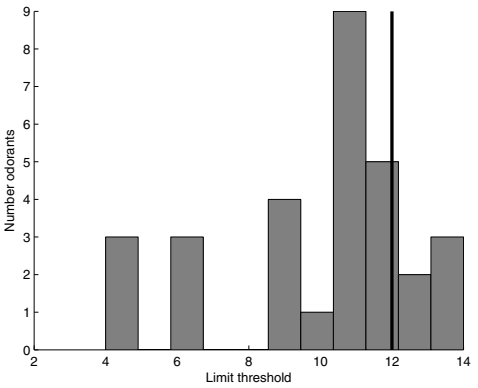


Fig. 4. Example of distribution of limit thresholds for a hidden layer neuron and a set of 30 character-based odorants with noise. The value, 12, selected as limit threshold, θ_j^O, for this neuron allows the neuron only fires for 5 odorants which have a threshold, θ_j, ≥ 13. Therefore, if the selected percentage is high, the limit threshold will be high too and the neuron will fire for a few odorants, the neuron will be more selective. If the percentage is low, the selectivity of the neuron is also low.

2.6 Results

We have taken all possible integer percentages, 0-100, and calculated the threshold for each neuron in the hidden layer. We have obtained the minimum discrimination error for each percentage and the spike rate for this minimum.

To achieve this minimum discrimination error, for each percentage used in the hidden layer we take all possible integer percentages and calculate the threshold for each neuron in the output layer. We calculate the discrimination error for all possible combinations and take the minimum observed. This value is the minimum discrimination error for a percentage used in the hidden layer.

In this section, we divide the results of the comparison of the different types of thresholds (homogeneous and heterogeneous) in two parts. First, we show the results for different sets of odorants: 15 orthogonal odorants, 30 orthogonal odorants, 15 character-based odorants and 30 orthogonal odorants; and different connection probabilities for the hidden layer, p_c. Finally, we present the results for a particular case, which shows the relationship between discrimination error and spikes rate.

Different Sets of Odorants and Connection Probabilities. We have realized 10 simulations for each set of odorants. We have realized these simulations for different connection probabilities for the hidden layer (p_c), different Hebbian learning probabilities (p_+,p_-) and noise presence (absent or present). We have used low connection probabilities for p_c (0.1, 0.3, 0.5) based on studies that confirm this [6,9], a probability for p_w (0.5) because it is applied to a matrix with learning, and selected Hebbian learning probabilities which have been previously studied [8]. Using this Hebbian learning over 20 time steps.

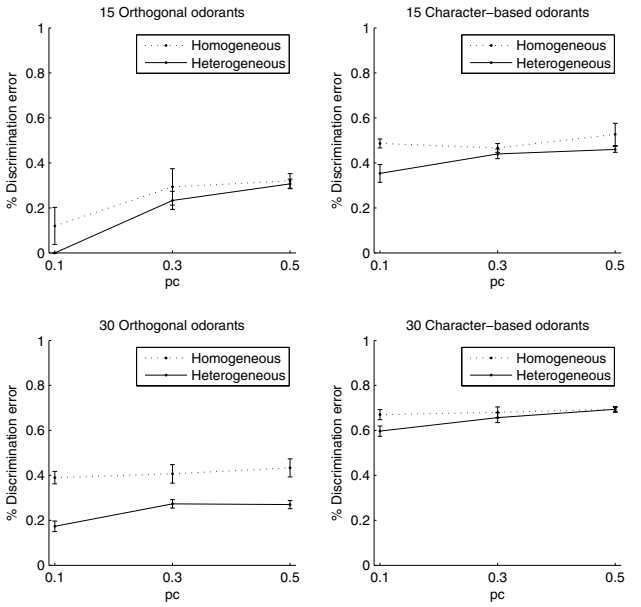


Fig. 5. Comparison of the different types of thresholds (homogeneous and heterogeneous) for different sets of odorants and connection probabilities with noise and the Hebbian learning probabilities which minimize the discrimination error. Sample means with 95% confidence intervals of standard errors (SE).

These averaged results (Fig.5) show that heterogeneous thresholds achieve lower discrimination errors. As they show how discrimination error increases with connection probability and number of odorants. Furthermore, by comparing these results with those obtained for odorants without noise (orthogonal odorants: 0.3% Homogeneous, 0.1% Heterogeneous; character-based odorants: 0.5% Homogeneous, 0.4% Heterogeneous), we can observe that the model is tolerant to noise and therefore it is robust.

High values for the discrimination error in the case of 30 character-based odorants can be explained by the similarity of some odorants, high noise which has been introduced in the input layer and network dimensions. Therefore, these

results suggest that there is some difficulty in discriminating for the selected network dimensions, especially for overlapping sets in the input.

Discrimination Error - Spike Rate. We have taken the averaged results, which we have seen above, and observed the relationship between discrimination error and spike rate for a particular case (Fig.6).

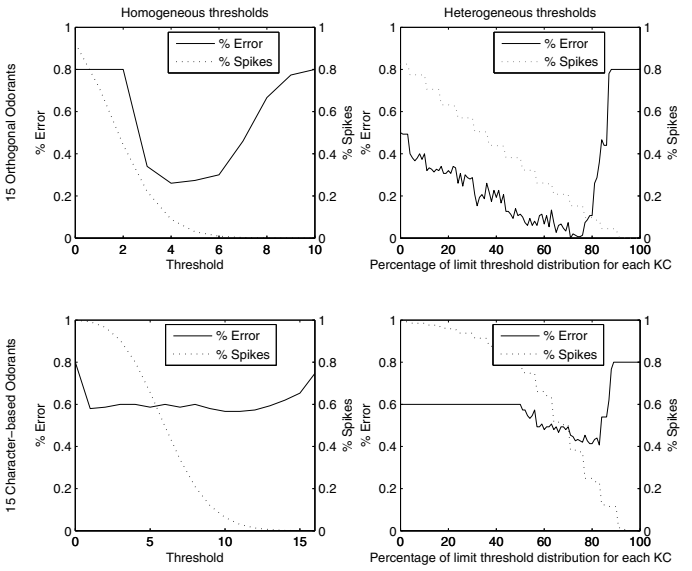


Fig. 6. Results for 15 orthogonal and 15 character-based odorants with noise and connectivity probability $p_c = 0.1$

These results show that minimum discrimination error is related to a low spike rate. This proves our hypothesis that high population sparseness in KC layer improves odorant discrimination.

The reason for this behavior is that if thresholds are too high, there will be very few neuron spikes in the hidden layer, and therefore odorant information which arrives to the output layer will be low, making discrimination impossible. However, if thresholds are too low, there will be a lot of neuron spikes in the hidden layer, and the output layer will have high population sparseness to make odorant discrimination possible.

2.7 Conclusions

We have shown that neural variability using heterogeneous thresholds improve odorant discrimination. Also, we have proven that the discrimination error is minimised when the sparse activity in the KC layer is increased. Finally, we have observed that discrimination error decreases for low connection probabilities between AL and MB.

Acknowledgments. This work was supported by the Spanish Government project TIN2010-19607 and R.H. acknowledges partial support by NIDCD-R01DC011422-01. Also, we thank Fernando Herrero-Carrón for useful discussions.

References

1. Angioy, A.M., Desogus, A., Barbarossa, I.T., Anderson, P., Hansson, B.S.: Extreme Sensitivity in an Olfactory System. Chemical Senses 28(4), 279–284 (2003)
2. Daly, K.C., Wright, G.A., Smith, B.H.: Molecular features of odorants systematically influence slow temporal responses across clusters of coordinated antennal lobe units in the moth Manduca sexta. J. Neurophysiol. (February 25, 2004)
3. Doleman, B.J., Lewis, N.S.: Comparison of odor detection thresholds and odor discriminablities of a conducting polymer composite electronic nose versus mammalian olfaction. Sensors and Actuators B 72, 41–50 (2001)
4. Dubnau, J., Grady, L., Kitamoto, T., Tully, T.: Disruption of neurotransmission in Drosophila mushroom body blocks retrieval but not acquisition of memory. Nature 411, 476–480 (2001)
5. Galan, R.F., Sachse, S., Galizia, C.G., Herz, A.V.: Odor-driven attractor dynamics in the antennal lobe allow for simple and rapid odor classification. Neural Computation 16(5), 999–1012 (2004)
6. García, M., Huerta, R.: Design Parameters of the Fan-Out Phase of Sensory Systems. Journal of Computational Neuroscience 15, 5–17 (2003)
7. Huerta, R., Nowotny, T., García, M., Abarbanel, H.D.I., Rabinovich, M.I.: Learning Classification in the Olfactory System of Insects. Neural Computation 16, 1601–1640 (2004)
8. Huerta, R., Nowotny, T.: Fast and Robust Learning by Reinforcement Signals: Explorations in the Insect Brain. Neural Computation 21, 2123–2151 (2009)
9. Jortner, R.A., Farivar, S.S., Laurent, G.: A simple connectivity scheme for sparse coding in an olfactory system. The Journal of Neuroscience (February 14, 2007)
10. Laurent, G., Stopfer, M., Friedrich, R.W., Rabinovich, M.I., Volkovskii, A., Abarbanel, H.D.I.: Odor encoding as an active, dynamical process: Experiments, Computation, and Theory. Neuroscience 24, 263–297 (2001)
11. Laurent, G.: Olfactory network dynamics and the coding of mutidimensional signals. Nat. Rev. Neurosci. 3(11), 884–895 (2002)
12. Marin, E.C., Jefferis, G.S., Komiyama, T., Zhu, H., Luo, L.: Representation of the glomerular olfactory map in the Drosophila brain. Cell 109(2), 243–255 (2002)
13. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in neurons activity. Bull. Math. Biophys. 5, 115–133 (1943)
14. Pérez-Orive, J.: Oscillations and Sparsening of Odor Representations in the Mushroom Body. Science 297, 359–365 (2002)
15. Tanaka, N.K., Awasaki, T., Shimada, T., Ito, K.: Integration of chemosensory pathways in the Drosophila second-order olfactory centers. Curr. Biol. 14(6), 449–457 (2004)
16. Turner, G.C., Bazhenov, M., Laurent, G.: Olfactory representations by drosophila mushroom body neurons. J. Neurophysiol. 99, 734–746 (2008)

Contrast Enhancement Mechanisms in the Retinothalamic Circuitry

Rubén Ferreiroa and Eduardo Sánchez

Grupo de Sistemas Intelixentes (GSI)
Centro Singular de Investigación en Tecnologias de la Información (CITIUS),
Universidade de Santiago de Compostela,
15782 Santiago de Compostela, Spain
rferreiroa@gmail.com, eduardo.sanchez.vila@usc.es
http://www-gsi.usc.es/index.html

Abstract. The center-surround organization of the receptive fields of retinal ganglion cells highlights the presence of contrast in visual stimuli. As the receptive fields of thalamic relay cells follow the same organization, it is assumed that nothing or little processing is carried out at the thalamic stage before the information reaches higher processing areas. However, recent data in cat showing that the number of thalamic relay cells doubles those of retinal ganglion cells opens the door to question how contrast information is kept in an enlarged representation of the visual stimulus at the thalamic stage. This paper is aimed at providing a plausible explanation by means of simulations performed with a realistic dynamic model of the retinothalamic circuit.

1 Introduction

The visual thalamus receives afferent input from retinal ganglion cells and projects its output to the visual cortex. In cat, the retinal input is mostly constituted by X and Y ganglion cells, which are the origin of the X and Y pathways, respectively. These pathways are functionally different and remain relatively well segregated through the dorsal lateral geniculate nucleus (dLGN) of the thalamus, until reaching the cortex (Cleland and Lee 1985; Cleland et al. 1971a; Hamos et al. 1987; Mastronarde 1992; Sincich et al. 2007; Usrey et al.1999). In this paper we are focused on the functional role of the retinothalamic X pathway, and concretely on the circuit of the dLGN layer A. This circuit is made up with relay cells and interneurons that takes as input the signals generated by the X ganglion cells. LGN relay cells are the main type of neurons and present a complex afferent input. They receive retinal as well as non-retinal afferents, the later coming from both local (mostly GABAergic interneurons and Perigeniculate cells) and external sources (feedback projection from cortical layer 6 and ascending projections from various scattered cell groups in the brainstem reticular formation). As in the retina, LGN relay cells present an excitatory center-inhibitory surround organization of their receptive field. Moreover, they can be classified either as ON- or OFF-center cells on the basis of their response to light stimuli.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 26–36, 2013.

The former are excited when stimulated by light in the center of their receptive fields, the later showing the opposite response. The excitatory center receives afferents from the retinal ganglion cells of the same sign (ON or OFF), while the inhibitory surround comes from local interneurons receiving input from retinal ganglion cells of the opposite sign.

The picture of the circuitry of the retinothalamic X pathway is incomplete without considering its topological features. The fact that the number of relay cells in the dLGN of the cat doubles the number of retinal ganglion cells while the number of local interneurons being half the number of retinal ganglion cells [5], posed the question about how relay cells are connected with both retinal ganglion cells and local interneurons. This was addressed by different authors [14], who carried out whole cell recordings in the cat's dLGN while stimulating with sparse-noise protocol. These observations have clarified the picture of the analysis performed by the center of the thalamic receptive fields but not by their surrounds, which is only mildly engaged by small stimuli. The contribution of the processing of the thalamic RF was explored in a topologically realist model of the dLGN [9], which is inspired by previous work by Molano et al. (2009) and Ringach (2007). Resorting on data of retinothalamic divergence available in the literature, Ferreiroa et al. (2013) found that the inhibitory surround coming from local interneurons being excited by the OFF-center ganglion cells improves the efficacy of the detection of stimulus edges carried out by the ON-center LGN cells. In addition, the model shows that tonic activity seems more suitable than bursting activity to salient fine details of the visual input.

2 Hypothesis

The goal of this study is therefore to bring light to the function of the dLGN circuit and its role on the retinothalamic X pathway. If the center-surround RF of a retinal ganglion cell can detect the contrast in light intensity between the center and its surround, we might wonder about the contribution of the center-surround organization at the thalamic stage. The first issue comes from the fact that at this stage there seems to be an enlargement of the visual stimulus, i.e the number of relay cells per ganglion cells follows a 2:1 ratio. In the field of image processing, such transformation could be accomplished by means of different types of interpolation schemes. However, interpolation is a smoothing operation that reduces the contrast of the original images. Therefore, the second issue regards with the mechanisms of the dLGN circuit to recover or mitigate the impact of such smoothing operation. Our working hypothesis states that the contribution of the inhibition evoked by the OFF center retinal ganglion cells on the ON-center LGN relay cells, is the key mechanism to compensate the loss of contrast during the retinothalamic image trasformation.

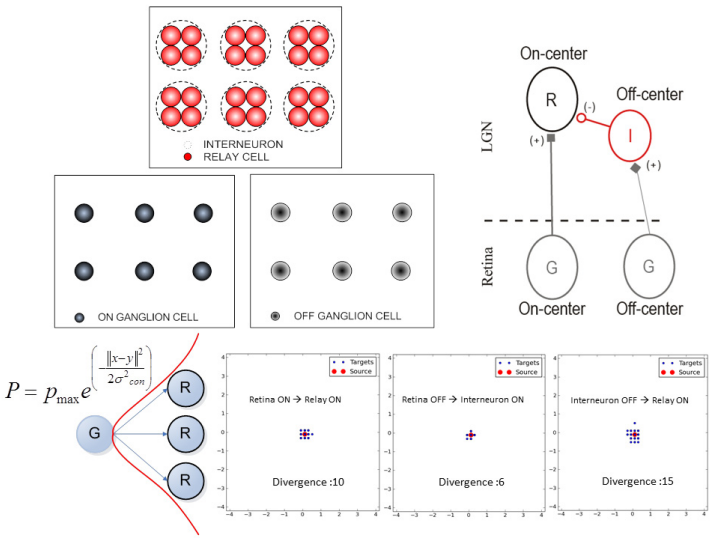


Fig. 1. Topology and spatial connectivity of the LGN model. Arrangement of relay cells and interneurons in LGN (upper left inset). Those nodes are the basic elements of the retinothalamic circuit, which is constituted by the ON and OFF channels (upper right inset). The connections between retinal ganglion cells and thalamic relay cells follow a Gaussian distribution probability (lower left inset). The degree of divergent connections between layers is derived from empirical studies (lower right inset).

3 Computational Methods

3.1 Circuit Layers

The retina and the thalamus are modeled by means of four different layers representing a patch parafoveal visual field of 8°x8°. The retina is simulated by two rectangular layers of 80x80 pixels, the first one representing the ON type ganglion cells, the second one the OFF type ganglion cells. The lateral geniculate nucleus is modeled by two rectangular layers, one for the relay cells and other for interneurons, with dimensions 160x160 and 40x40 pixels respectively. The ratio of relay cells per ganglion cells is 2:1, whereas the ratio of interneurons per neighboring relay cells is 1:4 (Molano and Martinez, 2009) (Fig. 1).

3.2 Coordinate System and Topology

The retina and the thalamus simulate a patch parafoveal visual field of 8°x8° mapped directly onto a lattice of size 80x80 pixels. Each position in the layer corresponds to a node and each node can define one type of neuron. In our case, nodes of the two layers of the thalamus correspond with one relay cells and one interneuron.

3.3 Connectivity

Each relay cell in the thalamus is first connected to its nearest neighbor in the retinal lattice, i.e. the one from which its polarity is inherited. The probability of each thalamic cell, re-centered at retinal coordinates at the x position of its first retinal input, is connected to another retinal ganglion cell located at the y position as a Gaussian function of their relative distance (x-y) [1]. The synaptic strength of the connections is also assumed to be a Gaussian function of the distance between the receptive-field centers [1]. The function for both connection probability and strength is as follows (see Fig. 1):

$$P = p_{max} e^{-(x-y)^2/2\sigma_{con}^2} \tag{1}$$

3.4 Input and Output Models

The stimuli used in our simulations are made up with a bright and a dark band connected with an abrupt (step function) transition (Fig. 2). Each stimulus is represented mathematically as a bi-dimensional 80x80 matrix, the matrix coefficients indicating the degree of intensity per point. The spatial organization of retinal ganglion cell receptive field is well captured by a difference of Gaussians model (Rodieck, Enroth-Cugell and Robson, 1966) in which the spatial receptive field is expressed as:

$$D(x,y) = \left(\frac{A}{2\pi\sigma_{cen}^2} e^{-(x-y)^2/2\sigma_{cen}^2} - \frac{B}{2\pi\sigma_{sur}^2} e^{-(x-y)^2/2\sigma_{sur}^2}\right) \tag{2}$$

The first Gaussian function describes the excitatory center and the second one the inhibitory surround. The sizes of the central and surround region is determined by the parameters σ_{cen} and σ_{sur}. Receptive field parameters for the ganglion cells have been set in accordance with Allen and Freeman [2].

The rate of the ganglion cells evoked by the visual stimulus is computed with a 2D convolution function:

$$r = r_0 + \int\int D(x,y)s(x,y)dxdy \tag{3}$$

where r represents the firing rate evoked by a stimulus $s(x,y)$, r_0 is the background firing, and $D(x,y)$ the difference-of-Gaussians function representing the receptive field of ganglion cells. The output of the convolution function for the ON and OFF ganglion cells is illustrated in Figure 2.

Finally, the stochastic nature of the ganglion cell activity is modeled through a pulse train generator that follows the statistics of a homogeneous Poisson process:

$$P_{T(n)} = \frac{rT}{n!} e^{-rT} \tag{4}$$

Fig. 3 shows the steps involved in the simulation of the neural activity at the different stages of our retinothalamic model.

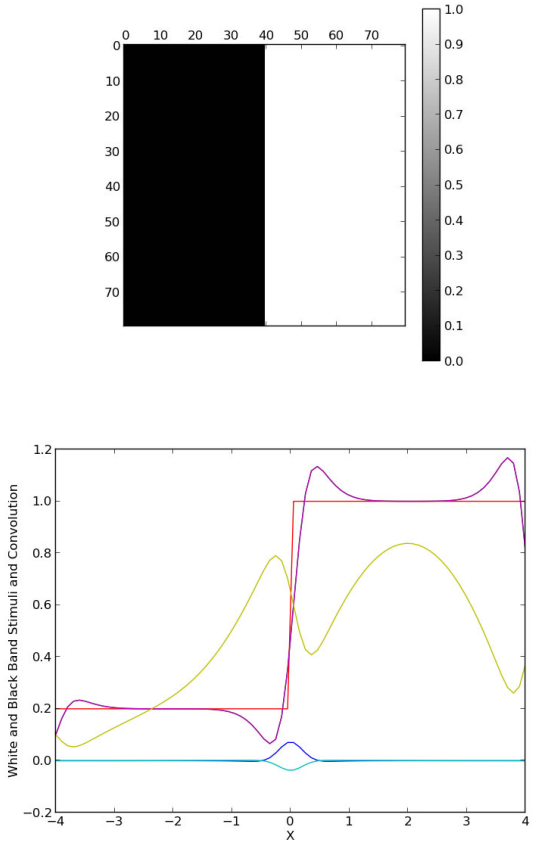


Fig. 2. Band Stimulus and retinal output. Band Stimuli with different degrees of luminance are used to analyze the functional properties of the circuit (top). Retinal output for ON (upper right line) and OFF cells (upper left line) are generated after convolving the center-surround receptive fields with the band stimulus (bottom).

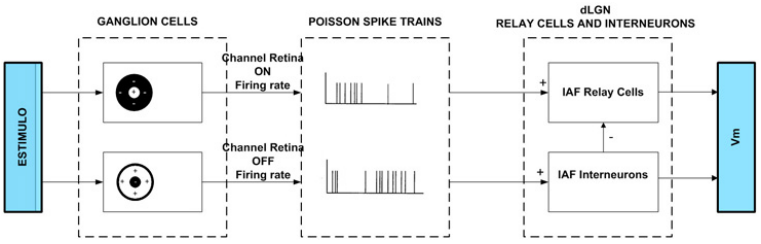


Fig. 3. General overview of the dLGN model

3.5 Neuron Model

We use an implementation of Eugene M. Izhikevich model [3], as implemented in the NEST simulator [4], for the thalamocortical cells and interneurons of dorsal lateral nucleus in the cat. This model reproduces regular spiking, fast spiking and bursting behavior selecting the appropriate parameters in the following equations.

$$C_m \frac{dV}{dt} = -k(V - V_r) + (V - V_t) - U + I \tag{5}$$

$$\frac{dU}{dt} = a(b(V - V_r) - U) \tag{6}$$

where C_m is the capacitance, V is the membrane potencial, V_r is the resting membrane potential, V_t is the instantaneous threshold potential, U is the recovery current, a is the recovery time constant, b a parameter that determines whether U is an amplifying ($b \leq 0$) or a resonant ($b \geq 0$) factor, and I the injected DC current. The dynamics of spike generation is governed by the following equation:

$$V \geq V_{peak} \begin{cases} V = c \\ U = U + d \end{cases} \tag{7}$$

where V_{peak} is the spike cutoff value, c the voltage reset value, and d describes the total amount of outward minus inward currents activated during the spike and affecting the after-spike behavior.

All these parameters can easily be fitted to simulate both thalamocortical relay cells and interneurons.

3.6 Contrast Definition

In visual perception, contrast is determined as the difference between the light intensities of two different regions of the visual field. To compute and compare the contrast information associated to the output of the different layers of the retinothalamic circuitry, we use the K contrast factor. This factor is defined as:

$$K = \frac{L_H - L_L}{L_H} \; 0 \leq K_m \leq 1 \tag{8}$$

where L_H represents either higher luminance or spike count, and L_L either lower luminance or spike count for each image representation (stimulus, retinal output, and LGN output) of the circuitry.

4 Results

4.1 Retina ON Channel

Figure 4 show the output of ON-center ganglion cells of the retina in terms of the number of spikes generated during a 70 ms interval after the stimulus

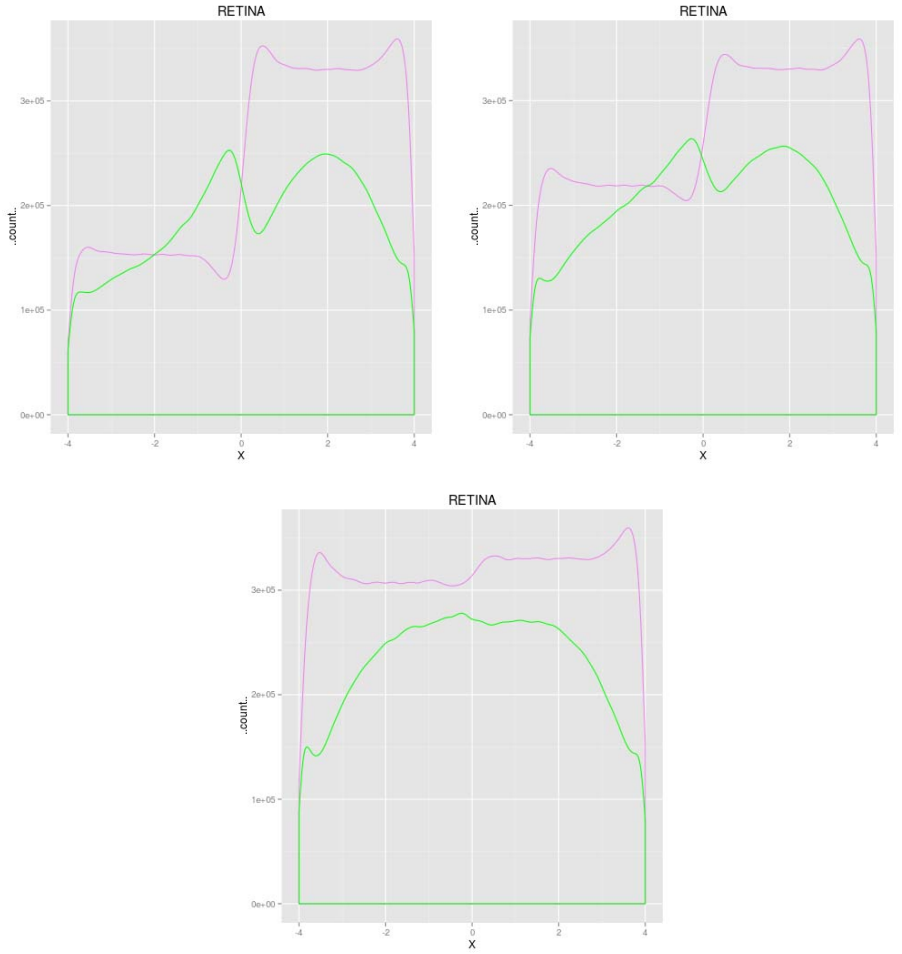


Fig. 4. Retina output after the presentation of stimuli with varying K contrast factor: 0.8 (top left), 0.5 (top right), and 0.1 (bottom)

presentation. Three different white-grey band stimuli were used, the white region kept constant in all three cases, while the grey region being increased on its light intensity values (0.2, 0.5 and 0.9). This means a K contrast factor (see Methods) of 0.8, 0.5 and 0.1, respectively.

The violet and green line in Figure 4 corresponds to the output of the ON and OFF retinal ganglion cells. It is interesting to point out that as the light intensity increases on the grey region, the spike count also increases while keeping almost constant the activity on the white region.

The simulations show that the center-surround receptive field of the retinal ganglion cells highlights the edge of the stimuli. The positive peak at the edge of

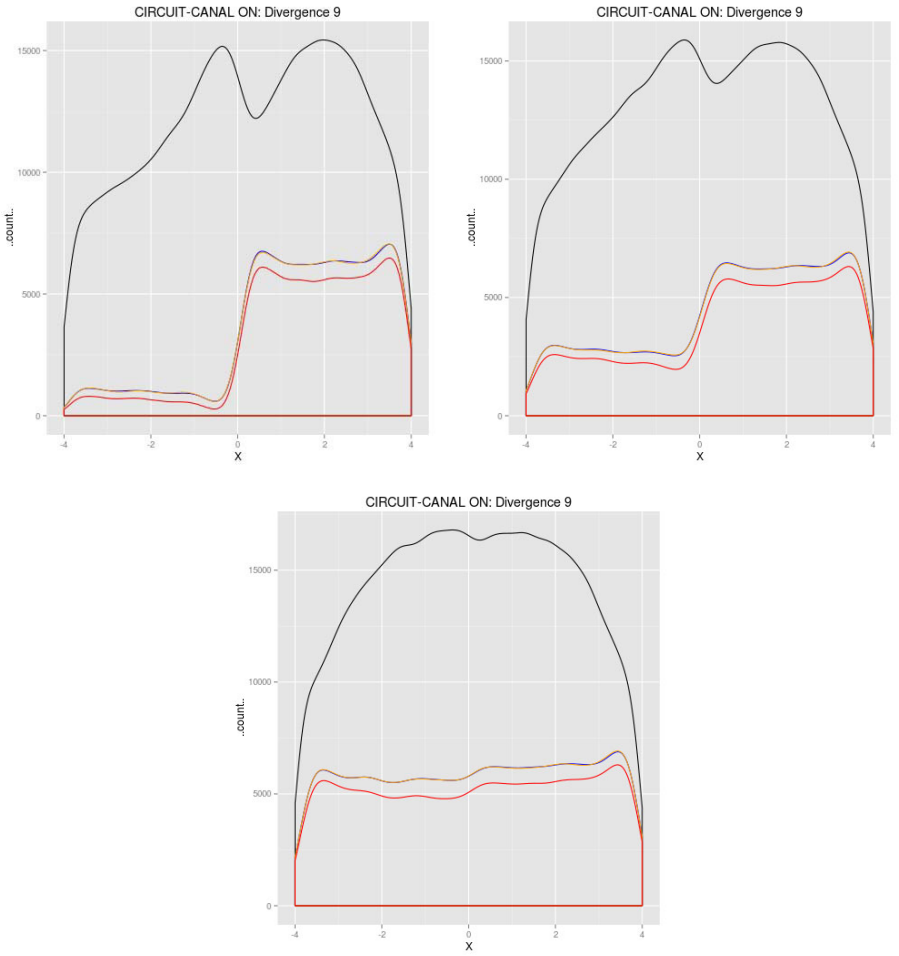


Fig. 5. LGN Output during the 0-70 ms interval. The LGN output is represented as the number of spikes generated on the LGN relay cells. LGN ON channel output (orange line), LGN OFF channel local interneurons output (black line), and LGN complete circuit output (blue and red lines) are shown after presenting three different stimulus with K contrast factor: 0.8 (top left), 0.5 (top right), and 0.1 (bottom).

the white band, located in the (0,4) interval of Fig. 4, and the negative peak at the edge of the dark grey band, located in the (-4,0) interval of Fig. 4, enhance the perceived contrast of the real stimuli.

4.2 LGN ON Channel

In the following experiment we present the output of the ON-center LGN relay cells after being excited from their corresponding ON-center retinal ganglion

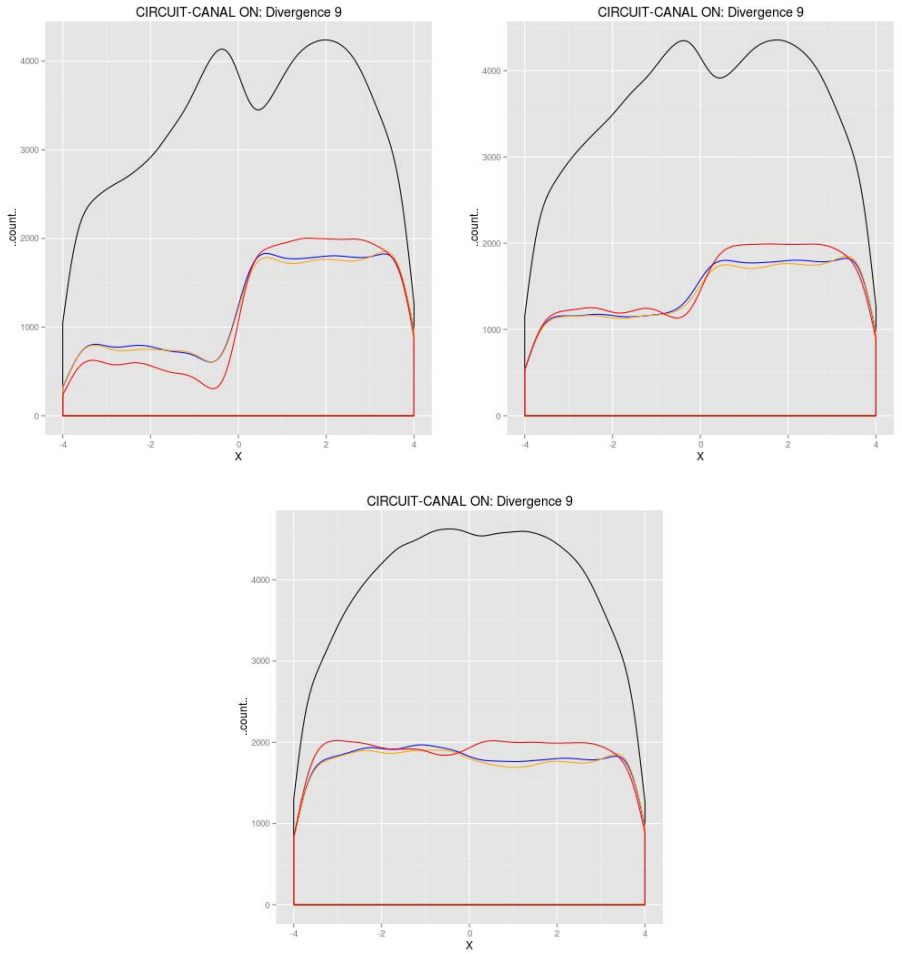


Fig. 6. LGN Output during the 50-70 ms interval. LGN ON channel output (orange line), LGN OFF channel local interneurons output (black line), and LGN complete circuit output (blue and red lines) are shown after presenting three stimulus with K contrast factor: 0.8 (top left), 0.5 (top right), and 0.1 (bottom). The blue line represents a divergence of 1 between local interneurons and LGN ON-center relay cells, while the red line represents a divergence of 31.

cells. All the results are obtained with a degree of divergence of 9 between On-center retinal ganglion cells and On-center LGN relay cells. This means one ganglion cell connected with nine thalamic relay cells of the same sign (see methods).

The orange line in Figure 5 represents the output of the LGN ON channel. The output of the LGN ON channel represents the transition from a layer of size 80x80, corresponding with the retina ON channel, to a greater layer of

Table 1. K contrast factor of the visual stimulus and the output of the retina ON channel (0-70 ms interval)

$K_{stimulus}$	K_{retina}
0.8	0.63
0.5	0.41
0.1	0.08

Table 2. K contrast factor at different time scales for LGN ON channel and LGN complete circuit

0-70 ms	50-70 ms
0.8 $K_{channelon}$=0.92,$K_{circuit}$=0.95	$K_{channelon}$=0.66,$K_{circuit}$=0.85
0.5 $K_{channelon}$=0.61,$K_{circuit}$=0.65	$K_{channelon}$=0.34,$K_{circuit}$=0.44
0.1 $K_{channelon}$=0.12,$K_{circuit}$=0.14	$K_{channelon}$=0.10,$K_{circuit}$=0.09

size 160x160. It is observed here how the thalamic ON channel is performing a smoothing function that causes a loss of sharpness and decrease in contrast at the edges of the input upcoming from the retina ON channel.

4.3 LGN Complete Circuit

Starting from the configuration of the section 4.2, we now activate the local interneurons of channel OFF (see Figure 1) keeping constant the degree of divergence between Retina OFF and Interneurons to 4, and changing the divergence between interneurons and LGN ON-center relay cells to 1 (blue line, see Figures 5 and 6) and 31 (red line, see Figures 5 and 6). The goal is to study the inhibitory effect coming from the OFF channel over the ON Channel at different time intervals.

As it is observed from the LGN OFF channel output (Figure 6), the inhibition is stronger over the negative peak of the grey band of the upcoming input, and weaker over the positive peak of the white band. The overall effect is to increase the difference between the positive and negative peaks, being this effect stronger for the circuit with divergence 31 during the 50-70 ms interval (red line, see Figure 6).

Tables 1 and 2 show the values of the K contrast factor for the retina output, the LGN ON channel output, and the LGN complete circuit output. Three interesting aspects are derived from these data: (1) contrast values of the LGN ON channel output are lower than those obtained from the retina output for the 50-70 ms interval, (2) contrast values of the LGN complete circuitry are higher than those obtained from both the retina output and LGN ON channel output for the 50-70 ms interval, and (3) contrast values for the 0-70 ms interval in both the LGN ON channel and complete circuitry are higher than those obtained from the retina output.

5 Discussion

The contrast values of Tables 1 and 2 need a detailed discussion. While points (1) and (2) presented in the last section support the working hypothesis of the paper, point (3) is quite controversial. The contrast values obtained when considering the 0-70 ms interval indicate that the LGN ON channel output improves the contrast information presented at the retinal output. This is an striking result when compared with the outcome of classical interpolation methods widely used in image processing software, and suggest that our initial hypothesis should be revised. As the results obtained within the 50-70 ms fully confirm what it was expected, we need to further analyze the impact of the selection of the temporal window as well as the temporal coding of the visual stimulus in the retinothalamic pathway.

References

1. Alonso, J.M., Yeh, C.I., Weng, C., Stoelzel, C.: Retinogeniculate connections: a balancing act between connection specificity and receptive field diversity. Progress in Brain Research 154 (2006)
2. Allen, E.A., Freeman, R.D.: Dynamic Spatial Processing Originates in Early Visual Pathways (2006)
3. Izhikevich, E.M.: Simple Model of Spiking Neurons. IEEE Transactions of Neural Networks 14(6) (November 2003)
4. Gewaltig, M.O., Diesmann, M.: Nest (neural simulation tool). Scholarpedia 2(4), 1430 (2007)
5. Madarász, M., Gerle, J., Hajdu, F., Somogyi, G., Tombol, T.: Quantitative histological studies on the lateral geniculate nucleus in the cat. II. Cell numbers and densities in the several layers. J. Hirnforsch. 19(2), 159–164 (1978)
6. Ringach, D.L.: Haphazard Wiring of Simple Receptive Fields and Orientation Columns in Visual Cortex. Journal of Neurophysiology 92, 468–476 (2007)
7. Yeh, C.I., Stoelzel, C.R., Weng, C., Alonso, J.M.: Functional Consequences of Neural Divergence Within the Retinogeniculate Pathway. J. Neurophysiol. 101 (2009)
8. McCormick, D.A., Huguenard, J.: A model of the electrophysiological properties of thalamocortical relay neurons. Journal of Neurophysiology 68, 1384–1400 (1992)
9. Ferreiroa, R., Sanchez, E.: Functional properties of a realistic model of dLGN. Neurocomputing (2013)
10. Palmer, S.E.: Vision Science. MIT (1999)
11. Dayan, P., Abbott, L.F.: Theorical Neuroscience. MIT (2001)
12. Rieke, F., Warland, D.: Spikes exploring the neural code. MIT (1999)
13. Koch, C.: Biophysics of Computation. Oxford University Press (1999)
14. Manzón, M.M., Rodríguez, M.M., Otero, L.M.: DEA. How the thalamus chages What the cat's eye tells the cat's brain (2009)

Neural Spike Activation in Hippocampal Cultures Using Hebbian Electrical Stimulation

V. Lorente[1], José Manuel Ferrández-Vicente[1,3], Fco. Javier Garrigós-Guerrero[1], Félix de la Paz López[2], José Manuel Cuadra-Troncoso[2], José Ramón Alvarez-Sanchez[2], and Eduardo Fernández[3]

[1] Departamento de Electrónica, Tecnología de Computadores y Proyectos, Universidad Politécnica de Cartagena, Spain
[2] Dpto. de Inteligencia Artificial - UNED - Madrid (Spain), Spain
[3] Instituto de Bioingeniería, Universidad Miguel Hernández, Alicante, Spain
victor.lorente@upct.es

Abstract. Electric stimulation has been widely used to induce changes in neuronal cultures coupled to microelectrode arrays (MEAs). In this paper, we used different electrical stimulation protocols on dissociated cultures of hippocampal cells for studying the electrical properties of the process. We show that persistent and synchronous stimulation of adjacent electrodes may be used for creating adjacent physical or logical connections in the connectivity graph following Hebb's Law modifying the neural responses principal parameters.

Keywords: Cultured neural network, Hebbian Law, induced plasticity, learning.

1 Introduction

Biological brains use millions of biological processors, with dynamic structure, slow commutations compared with silicon circuits [1,9], with low power consumption and unsupervised learning. The use of dissociated cortical neurons cultured onto MEAs represents a useful experimental model to characterize both the spontaneous behavior of neuronal populations and their activity in response to electrical and pharmacological changes. Learning is a natural process that needs the creation and modulation of sets of associations between stimuli and responses. Many different stimulation protocols have been used to induced changes in the electrophysiological activity of neural cultures looking for achieve learning [8,4] and low-frequency stimulation has brought good results to researchers enhancing bursting activity in cortical cultures [10,11]. Hebbian learning describes a basic mechanism for synaptic plasticity wherein an increase in synaptic efficacy arises from the presynaptic cell's repeated and persistent stimulation of the postsynaptic cell. The theory is commonly evoked to explain some types of associative learning in which simultaneous activation of cells leads to pronounced increases in synaptic strength. Basically the efficiency of a synaptic connection is

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 37–47, 2013.

increased when presynaptic activity is synchronous with post-synaptic activity. In this work, we use this kind of stimulation to create adjacent physical or logical connections in the connectivity graphs using Hebb's Law. In previous papers, we used a specific low-frequency current stimulation on dissociated cultures of hippocampal cells to study how neuronal cultures could be trained with this kind of stimulation [6,5]. We showed that persistent and synchronous stimulation of adjacent electrodes may be used for creating adjacent physical or logical connections in the connectivity graph following Hebb's Law. In later experiments, we have used different parameters for this stimulation to check if those connections can be created stimulating with different configurations. The results provided in this paper show that low-frequency stimulation can create adjacent connections with different amplitude values modifying the electrical characteristics of the acquired neural responses. The outline of the paper is as follows. We present the methods for addressing Hebbian Learning throw electrical stimulation. The following section shows the results obtained using a specific stimulation with our experimental setup on hippocampal cultures to train them. We conclude by discussing some crucial aspects of the research and the remaining challenges.

2 Methods

2.1 Cell Culture Preparation

Dissociated cultures of hippocampal CA1-CA3 neurons were prepared from E17.5 sibling embryos (Figure 1). During the extraction of the hippocampus a small amount of cortical tissue will have inevitably also been included. Tissue was kept

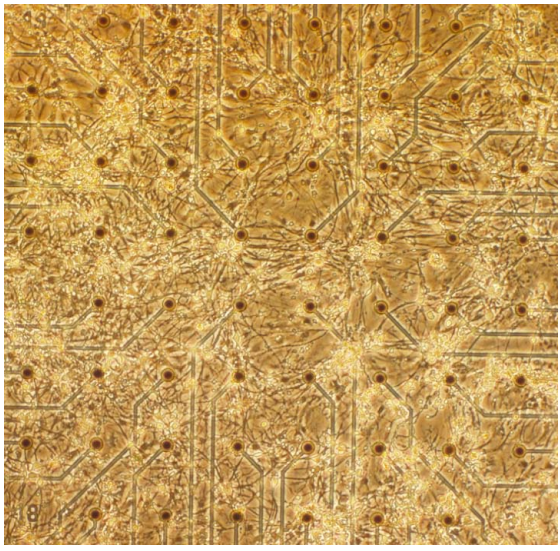


Fig. 1. Hippocampal CA1-CA3 culture (21 DIV) on a microelectrodes array

in 2ml of HBSS. 10mg/ml of trypsin was added to the medium and placed in a 37° C water bath for 13 min for subsequent dissociation. The tissue was then transferred to a 15 ml falcon containing 4ml of NB/FBS and triturated using combination of fine pore fire polished Pasteur pipettes (Volac). Cells were then transferred onto 12 well plates (Corning Incorporated) containing glass coverslips (Thermo Scientific).

The coverslips were pre-treated overnight with PDL (50mg/ml), a synthetic molecule used as a coating to enhance cell attachment. The PDL was then aspirated away and the coverslips washed twice with PBS. This was then followed by a final coating of laminin (50μg/ml), a protein found in the extracellular matrix, to further help anchor the dissociated hippocampal cells. The cells were maintained in a mixture of 500ml NB/B27 (promotes neural growth) and 500ml NB/FBS (promotes glial growth), each supplemented with Glutamax and Pen/Strep (dilution 1/100). Glutamax improves cells viability and growth while preventing build up of ammonia and Pen/Strep helps to prevent any infections. Cell density for each coverslip was roughly 200000 cells. Cells were kept in an incubator at 37 ° C in 6% CO2.

2.2 Experimental Setup

Microelectrode arrays (Multichannel systems, MCS) consisted of 60 TiN/SiN planar round electrodes (200 μm electrode spacing, 30 μm electrode diameter) arrange in a 8x8 grid were used. The activity of all cultures was recorded using a MEA60 System (MCS). After 1200X amplification, signals were sampled at 10kHz and acquired through the data acquisition card and MCRack software (MCS). Electrical stimuli were delivered through a two-channel stimulator (MCS STG1002) to each pair of electrodes.

2.3 Experimental Protocol

A total of 24 cultures were used in five experiments of 2-3 weeks duration. In every experiment 4-5 cultures were stimulated with a specific electrical stimulation protocol. In experiments E1 to E3 a low frequency current stimulation with different parameters for each experiment was used. Experiments E4 to E5 used a more aggressive stimulation called Tetanization. Experiments were started when neural cultures had 14 Days in Vitro (DIV) and were carried out during 2-3 weeks. Table 1 summarizes the experiments and the stimulation parameters applied to the cultures.

In experiment1 (E1), cultures 48 to 52 were stimulated during 16 days with trains of 5 biphasic pulses cathodic-first (50 μA peak, 100 μs phase, 50ms Interval), that were delivered every 3s for 10 minutes. In experiment2 (E2), cultures 68 to 72 were stimulated during 10 days with trains of 5 biphasic pulses cathodic-first (60 μA peak, 100 μs phase, 50ms Interval), that were delivered every 3s for 8 minutes. Experiment3 (E3), cultures 73-77 were stimulated during 11 days with trains of 5 biphasic pulses cathodic-first (40 μA peak, 100 μs phase, 50ms Interval), that were delivered every 3s for 8 minutes.

Table 1. Experiments, stimulation durations in days, cultures identifiers and stimulation parameters

Experiment	Duration	Cultures ID	Stimulation Parameters
E1	16	48-52	Low freq.: 5x(50 μA, 100, -50, 100, 20Hz), 3s ITI, 10'
E2	10	68-72	Low freq.: 5x($\pm$60 μA, 20Hz), 3s ITI, 8'
E3	11	73-77	Low freq.: 5x($\pm$40 μA, 20Hz), 3s ITI, 8'
E4	10	78-82	Tetani: 20x(600mV, 400μs, -600mV, 400μs, 50ms IPI), 6s ITI, 10'
E5	13	83-86	Tetani: 20x(400mV, 400μs, -400mV, 400μs, 50ms IPI), 6s ITI, 10'

Tetanization stimulation was applied for experiments 4 and 5. In experiment4 (E4), cultures 78 to 82 were stimulated during 10 days with trains of 20 biphasic pulses cathodic-first (600 mV peak, 400 μs phase, 50ms Interval), that were delivered every 3s for 10 minutes. In experiment5 (E5), cultures 73 to 76 were stimulated during 13 days with trains of 20 biphasic pulses cathodic-first (400 mV peak, 400 μs phase, 50ms Interval), that were delivered every 3s for 10 minutes.

In every experiment, the electrophysiological activity of the cultures was previously analyzed and connectivity diagrams based on cross-correlation were obtained for each culture. Two pairs of electrodes with an acceptable spiking activity and no logical connections between them were selected for stimulation. The following days every stimulation session would follow these steps:

1. Spontaneous activity was recorded for 2 min after a recovery period.
2. Cultures were then stimulated through the two pairs of electrodes using the corresponding stimulation protocol.
3. Spontaneous activity was recorded for 2 min after the stimulation.

2.4 Analysis Performed

We observed the spontaneous activity of the cultures before and after the stimulation experiments, as well as their evoked response to the applied stimulus. Extensive burst analysis, post-stimulus time histograms and functional connectivity were the main analysis performed to the registered data. Functional connectivity [2,3] captures patterns of deviations from statistical independence between distributed neurons units, measuring their correlation/covariance, spectral coherence or phase locking. Functional connectivity is often evaluated among all the elements of a system, regardless whether these elements are connected by direct structural links; moreover, it is highly time-dependent (hundreds of milliseconds) and model-free, and it measures statistical interdependence (e.g. mutual information) without explicit reference to causal effects. Correlation and information

theory-based methods are used to estimate the functional connectivity of in-vitro neural networks: Cross-correlation, Mutual Information, Transfer Entropy and Joint Entropy. Such methods need to be applied to each possible pair of electrodes, which shows spontaneous electrophysiological activity. For each pair of neurons, the connectivity method provides an estimation of the connection strength (one for each direction). The connection strength is supposed to be proportional to the value yielded by the method. Thus, each method is associated to a matrix, the Connectivity Matrix (CM), whose elements (X, Y) correspond to the estimated connection strength between neuron X and Y. High and low values in the CM are expected to correspond to strong and weak connections. By using such approach, inhibitory connections could not be detected because they would be mixed with small connection values. However, non-zero CM values were also obtained when no apparent causal effects were evident, or no direct connections were present among the considered neurons. In our experiments, Connectivity maps offered a visualization of the connectivity changes that occur in the culture. Connectivity maps were generated using the connectivity matrix (CM) obtained after applying the analysis and Cross-Correlation or Mutual Information. By setting thresholds in the CM, it is possible to filter out some small values that may correspond to noise or very weak connections. In consequence, these maps show the strongest synaptic pathways, and can be used for visualizing the neural weights dynamics, and validate the achieved learning.

Spike parameters, like height and width of positive and negative voltages, as the evolution of the spiking activity were also subject of research. We were very interested in analysing how the different stimulations could affect the electrical properties of the cultures.

3 Results

Low-frequency current stimulation and tetanic stimulation had both an impact on the electrophysiological responses of the cultures, as previous studies had reported [4,7]. Raster plots showed that all of the stimulations provided induce changes in the firing frequency of the cultures. We can observe some kind of reorganization in the firing activity, from a nearly random continuous spiking activity to a more discrete bursting spiking activity. After the third week in vitro, this bursting activity becomes more frequent and robust and this effect is much more evident than during the first weeks (Figure 2). First experiments showed that this effect takes effect initially with low-frequency stimulation, however from our last experiments it may be concluded that both stimulations have a frequency impact on the spiking activity of the culture. It may be concluded that this effect is more evident using low-frequency neural stimulation.

The change on the spiking activity of the cultures can also be seen analytically by observing the instantaneous firing frequencies (Figure 3) and the interspike intervals of the neural cultures. Instantaneous firing frequency graphs shows that stimulated electrodes start firing in more separated period of times after stimulation but each firing period last longer. In addition, interspike intervals

Fig. 2. Raster plots extracted from cultures of experiments E2 and E5. (a) (21DIV) and (b) (32DIV) belong to ID68 from E2, (c) (23DIV) and (d) (30DIV) belong to ID86 from E5. Each figure is divided in two graphs, which show the spiking activity of the culture before and after stimulation. Raster plots show a change in the spiking activity, changing from a uniform activity before stimulation to a more concentrated activity after stimulation. This result is emphasized after the third week in vitro due to maturing occurred in the cultures.

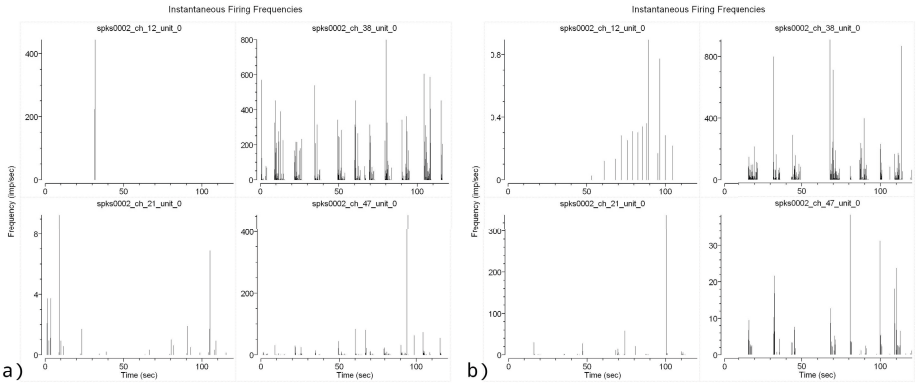


Fig. 3. Instantaneous firing frequencies of stimulated electrodes 12-21 and 38-47 in the culture ID68 (22DIV) from E2 before (a) and after (b) stimulation. A change in the spiking periods can clearly be seen, with less active frequencies and with longer periods. This culture only created a connection between the second pair of electrodes (38, 47), which had an impact on the instantaneous firing frequencies. Connected electrodes fire in the same firing periods, whereas not connected electrodes had no firing relation.

show the same results observed in the spiking periods but also it can be seen analytically the ISI decrease both in value and dispersion. Both effects are related to the neural stimulation, which modulates the firing capabilities of the cultures.

Connectivity diagrams based on cross-correlation between electrodes showed some kind of connections reorganization after stimulations, concentrating them in a few electrodes. Furthermore, adjacent physical or logical connections in the connectivity graph following Hebb's law appeared in some pairs of stimulated electrodes (Figure 4). Electrodes with created connections between them can distinctly be detected with the instantaneous firing frequencies graphs. Figure 3 showed two pair of stimulated electrodes (12, 21 and 38, 47) before and after the stimulation session. The firing periods of the electrodes from the second pair follow exactly each other, whereas the firing periods of the first pair of electrodes do not match. Furthermore, the electrodes of the second pair change both the firing periods after stimulation. This features indicates that there exits a strong connection between them.

Analysing spike parameters such as peaks heights and widths and number of spikes have lead us to an important result. Both types of stimulation, low-frequency and tetanic stimulation, produced a reactivation of neurons over time which lead to the creation of adjacent physical or logical connections in the connectivity graph following Hebb's Law. Figure 5 shows the mean values of spike parameters of stimulated electrodes for each register number. Figure 5a and b belong to the culture ID48 from E1, stimulated with low-frequency. We can observe in these figures that each electrode presented a good evolution of its parameters until register number 20. Negative peak height decreased, positive peak height increased, width of negative peak slightly increased and number

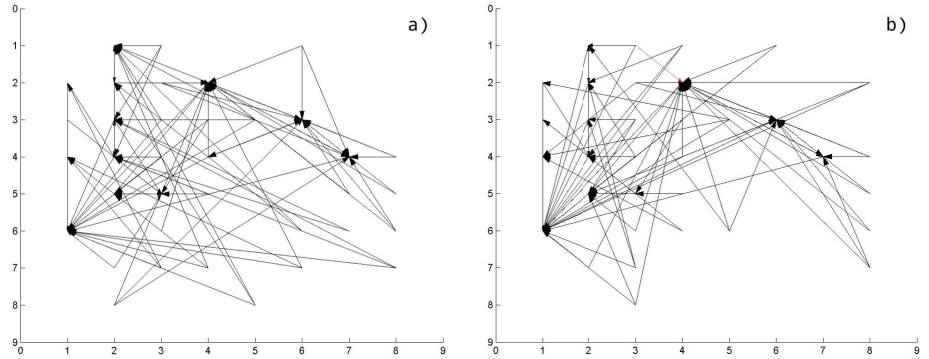


Fig. 4. Connectivity graphs based on cross-correlation between electrodes. The graph belong to the culture ID48 (E1) at 25 DIV. Pair of electrodes 31, 42 and 52, 53 were stimulated with low-frequency current stimulation with 50 μA biphasic pulses. (a) No logical connections were observed before stimulation. (b) A connection (red arrow) between electrodes 31 and 42 has appeared.

of spikes increased. This positive evolution of the parameters was due to the activation of more new neurons near the electrodes, which produced a higher value of the spikes parameters.

Fig. 5c and d belong to the culture ID86 from E5, stimulated with tetanic stimulation. The first electrode stimulated showed a positive evolution of its parameters until register number 10, whereas the second electrode became nearly stable until register number 12. In this case, the neural activation took place in the first electrodes, which lead to the creation of few connections over time. The changes in the spikes parameters can also be seen clearly in Figure 6, which shows the spikes waveform from cultures ID68 and ID86 at the stimulated electrode 38 and 62, respectively. The figure presents the spikes waveform at early stages pre-stimulation (blue) and late stages post-stimulation (green). The acquired neural spikes increased their number and become bigger and more consistent (less variant slope), due to the activation of more neural aggregates around the target-electrode with more mature and permanent electrical responses, which induces a more solid and synchronous activation related with paired electrode.

All E1 cultures created a connection between the paired stimulated electrodes, whereas 60% of the cultures of E3 and E4, and only 20% of the cultures of E2 and E5 showed that connection. In some cases, the connection was intermittent, lasting one to several days. In others, a persistent connection was created. Finally, some cultures did not create any kind of connections. In this way, Hebbian tetanization created ad-hoc permanent or transient logical connections by modifying the efficiency of the paths.

between the selected electrodes. We speculate that unconnected electrode cultures may be caused by a not-homogeneous culture growth between the electrodes or by the neurobiological properties of the connections as still need to be confirmed using histological techniques in future works. In this case, using

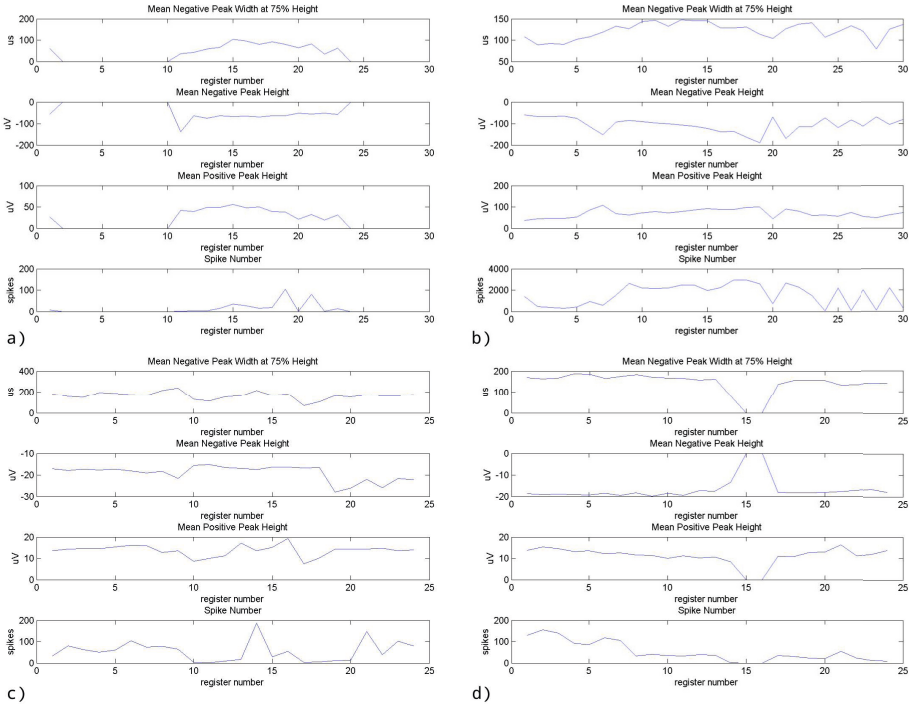


Fig. 5. For each electrode, Mean Negative Peak Height, Mean Positive Peak Height and Spike Number are represented. a) and b) graphs are for Electrode 31 and 42 from culture ID48. c) and d) graphs are from Electrode 62 and 72 from culture ID86.

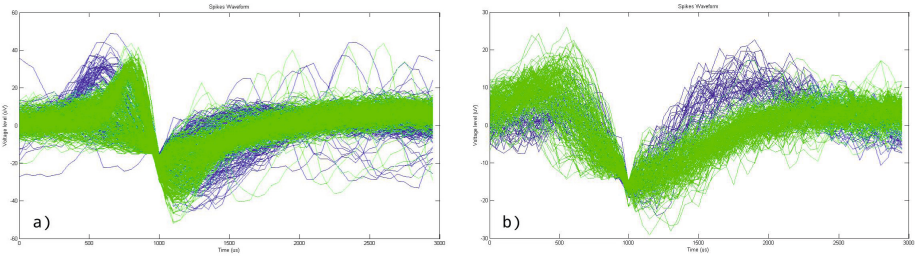


Fig. 6. Spikes waveform from cultures of ID68 and ID86 of experiments E2 and E5, respectively. (a) Spikes waveforms from electrode 38 at 21 DIV (blue) and 32 DIV (green). (b) Spikes waveforms from electrode 62 at 23 DIV (blue) and 30 DIV (green).

low-frequency current stimulation with 50 μA biphasic pulses provided the best results for creating connections following Hebb's law. Finally it can be concluded that low-frequency stimulation is better in terms of neural response stability, culture duration and connection-oriented processes than tetanization that evokes a more aggressive process in the hippocampal culture.

4 Conclusions

Learning in biological neural cultures is a challenging task. Different authors have proposed different methods for inducing a desired and controlled plasticity over the biological neural structure. Low-frequency stimulation and tetanization has brought good results to researchers enhancing bursting activity in cortical cultures. In this paper, we have shown that using these kind of stimulations it is possible to create adjacent physical or logical connections in the connectivity graph following Hebb's Law and such connections induce changes in the electrophysiological response of the cells in the culture, which can be observed in the different analysis performed. Furthermore, low-frequency stimulation induces permanent changes in most experiments using different values of current amplitude and stimulation patterns. Persistent and synchronous stimulation of relevant adjacent electrodes may be used for strengthen the efficiency of their connectivity graph. These processes may be used for imposing a desired behaviour over the network dynamics. In this work, different stimulation procedures are described in order to achieve the desired plasticity over the neural cultures, shaping in this way the functional connectivity of the neural culture by modifying their electrical neural responses acquired in the induced-connection process. In future works, we will try to refine what are the stimulation optimal parameters for inducing persistent changes in the cultured network. These induced connections will be used for driving a robot using Braitenberg's principles.

Acknowledgments. This work is being funded by grant 12361/FPI/09 from Séneca Foundation, Science and Technology Agency from the region of Murcia and by the project 2010V/PUNED/0011 from Universidad Nacional de Educación a Distancia.

References

1. Anderson, J.A., Rosenfeld, E.: Neurocomputing: Foundations of research. MIT Press (1988)
2. Bologna, L.L., Nieus, T., Tedesco, M., Chiappalone, M., Benfenati, F.: Low-frequency stimulation enhances burst activity in cortical cultures during development. Neuroscience 165, 692–704
3. Braitenberg, V.: Vehicles: experiments in synthetic psychology. MIT Press (1986)
4. Chao, Z.S., Bakkum, D.J., Potter, S.M.: Region-specific network plasticity in simulated and living cortical networks: comparison of the center of activity trajectory (cat) with other statistics. J. Neural. Eng. 4(3), 294–308 (2007)
5. Ide, A.N., Andruska, A., Boehler, M., Wheeler, B.C., Brewer, G.J.: Chronic network stimulation enhances evoked action potentials. J. Neural. Eng. 7(1) (February 2010)
6. Stegenga, J., Marani, E., Rutten, W.L.C.: Robustness of bursts in electrically stimulated neuronal networks. Journal of Neural Engineering

7. Jimbo, Y., Robinson, H.P., Kawana, A.: Strengthening of synchronized activity by tetanic stimulation in cortical cultures: application of planar electrode arrays. IEEE Trans. Biomed. Eng. 45(11), 1297–1304 (1998)
8. Landweber, L., Kari, L.: The evolution of cellular computing: Nature's solution to a computational problem. Biosystems 52(1/3), 3–13 (1999)
9. University of Heidelberg (2006)
10. Ruaro, M.E., Bonifazi, P., Torre, V.: Toward the neurocomputer: image processing and pattern recognition with neuronal cultures. IEEE Trans. Biomed. Eng. 52(3), 371–383 (2005)
11. Wagenaar, D.A., Pine, J., Potter, S.M.: Searching for plasticity in dissociated cortical cultures on multi-electrode arrays. Journal of negative results in biomedicine 52(16) (2006)

Thalamo-Cortical Network and Seizure Dynamics: A Computational Study

F. Vallone, S. Chillemi, M. Barbi, and A. Di Garbo

Istituto di Biofisica CNR, Sezione di Pisa,
via G. Moruzzi 1, 56124 Pisa, Italy
{fabio.vallone,santi.chillemi,michele.barbi,angelo.digarbo}@pi.ibf.cnr.it
http://www.pi.ibf.cnr.it

Abstract. Experimental data indicate that thalamic inputs are important factors for the generation and termination of seizures. In this paper a minimal biophysical model of cortico-thalamo-cortical network is investigated by a computational approach. The results show that a change in the amplitude of synaptic currents between thalamic and cortical neurons promotes seizure like dynamics. Moreover, the increase of the level of inhibition between neurons of the thalamic network is sufficient for seizure termination.

1 Introduction

About 50 million people worldwide have epilepsy, and they are usually controlled, but not cured, with medication. Although many studies have been made on seizures, the mechanisms of generation and termination still remain poorly understood(see for a complete review[1,2]). Recently a new vision of the epileptic seizures has been discovered[3]. Contrary to the traditional view, suggesting hypersynchronous neuronal activity during the ictal activity, an highly heterogeneous neuronal spiking activity was observed. In particular, seizure termination is described by a quasi-homogenous phenomenon leading to an almost complete cessation of spiking activity[3]. In addition it was found that the spike waveforms does not change at seizure termination, an indication that depolarization block is not the principal factor responsible for the cessation of spiking activity [3]. Obviously, the most important mechanisms, relevant for a deep understanding of seizure dynamics, are those driving the generation and termination of the ictal events. Among the possible mechanisms of generation and termination of seizure, thalamic inputs can play an important role. In fact the cortex is intimately connected with thalamus, and the cortico-thalamo-cortical excitatory loop can mediate network oscillations underlying epilepsies [4]. Moreover, in a recent experimental work it was shown that thalamocortical neuronal activity is required for post-stroke epilepsy; in addition a reduction of the activity of thalamocortical cells is sufficient to stop seizures [5]. Concerning the problem of seizure termination in [6] it was shown that a clear connection exists between extinction and spatial synchronization of populations. This general results could

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 48–61, 2013.

be useful to justify the possibility that the termination of seizures can arise from an emergent property of the network itself. Moreover, in a recent computational study it was shown that a depolarization block could be the primary factor for the seizure termination [7], but this result does not seem to be in agreement with the experimental data of Truccolo et al. [3]. Therefore, motivated by the above discussion the effects, of the synaptic connectivity of the cortico-thalamo-cortical network, on the dynamics of seizure generation and termination will be investigated computationally.

2 Methods

2.1 Model Description

The artificial network is composed by N_{PY} pyramidal neurons, N_{FS} FS interneurons, N_{RE} reticular neurons and N_{TC} thalamocortical neurons. A schematic representation of the network connectivity if reported in figure 1. The pyramidal neuron models are coupled by excitatory synapses and receive inhibitory inputs from the network of FS interneuron. For either the pyramidal neuron or the interneuron, a single compartment biophysical model is employed to describe its spiking activity. In particular, the adopted pyramidal and interneuron biophysical models were those proposed in [8]. The mathematical model of the $j-th$ pyramidal neuron reads:

$$C\frac{dV_j}{dt} = I_{P,j} - g_{Na}m_j^3h_j(V_j - V_{Na}) - g_K n_j^4(V_j - V_K) - g_M w_j(V_j - V_M) \\ -g_L(V_j - V_L) + I_{PP,j} + I_{IP,j} + I_{TP,j} + \eta_P\xi_{P,j}(t) \quad (1)$$

$$\frac{dm_j}{dt} = \alpha_{m,j}(1 - m_j) - \beta_{m,j}m_j \quad (2)$$

$$\frac{dh_j}{dt} = \alpha_{h,j}(1 - h_j) - \beta_{h,j}h_j, \quad (3)$$

$$\frac{dn_j}{dt} = \alpha_{n,j}(1 - n_j) - \beta_{n,j}n_j, \quad (4)$$

$$\frac{dw_j}{dt} = \frac{w_{j,\infty} - w_j}{\tau_{j,w}}, \quad (5)$$

where $C = 1\ \mu F/cm^2$, $I_{P,j} = I_P\,(j = 1, 2, ..N)$ is the external stimulation t. The maximal specific conductances and the reversal potentials are respectively: $g_{Na} = 100\ mS/cm^2$, $g_K = 80\ mS/cm^2$, $g_M = 1\ mS/cm^2$, $g_L = 0.15\ mS/cm^2$ and $V_{Na} = 50\ mV$, V_K = -100 mV, V_M = -100 mV, V_L = - 72 mV. The rate variables describing the currents are defined by: $\alpha_{m,j}(V_j) =$

$0.32(V_j+54)/[1-exp((V_j+54)/4)]$,$\beta_{m,j}(V_j) = 0.28(V_j+27)/[exp((V_j+27)/5)-1]$, $\alpha_{h,j}(V_j) = 0.128exp(-(V_j + 50)/18)$, $\beta_{h,j}(V_j) = 4/[1 + exp(-(V_j + 27)/5)]$, $\alpha_{n,j}(V_j) = 0.032(V_j + 52)/[1 - exp(-(V_j + 52)/5)]$, $\beta_{n,j}(V_j) = 0.5exp(-(V_j + 57)/40)$, $w_{j,\infty} = 1/[1+exp(-(V_j+35)/10)]$, $\tau_{j,w} = 400/[3.3exp((V_j+35)/20)+exp(-(V_j+35)/20)]$. In this model the onset of periodic firing occurs through an Hopf bifurcation for $I_P \cong 3.25\ \mu A/cm^2$ with a well defined frequency ($\nu \cong 5Hz$).

The current $I_{PP,j}$ arises from the excitatory coupling of the $j-th$ pyramidal neuron with the other cells, $I_{IP,j}$ describes the inhibitory current due to the coupling with the network of interneurons and $I_{TP,j}$ represents the excitatory inputs from TC cells. These currents will be defined in the next section.

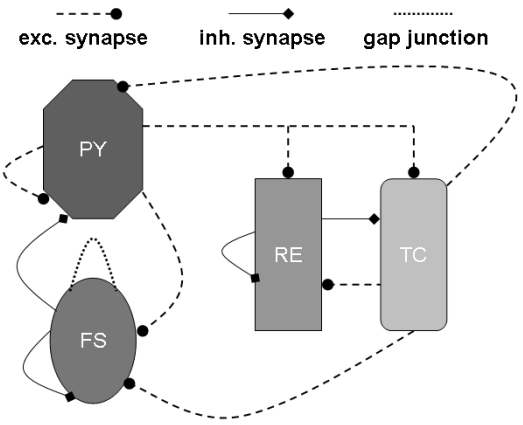


Fig. 1. Schematic representation of the neural networks connectivity. Pyramidal neurons (PY) receive excitatory inputs from thalamocortical cells (TC) and inhibitory inputs from FS interneurons. Thalamic reticular neurons (RE) receive excitatory inputs from pyramidal and TC neurons, and inhibit TC cells. FS interneurons are coupled by electrical synapses, inhibit PY neurons and receive excitatory inputs from PY on TC neurons.

To reproduce the membrane potential fluctuations each $j-th$ cell model is injected with the noisy current $\eta_P\xi_{P,j}(t)$, $\xi_{P,j}$ being an uncorrelated Gaussian random variable of zero mean and unit standard deviation $<\xi_{P,i},\xi_{P,j}>=\delta_{ij}, i \neq j = 1,2,3,,N_{PY}$. The adopted value of the parameter η_P was chosen to get realistic amplitude of the fluctuations of membrane potential.

The biophysical mathematical model of the $j-th$ FS interneuron reads:

$$C\frac{dV_j}{dt} = I_{F,j} - g_{Na}m_j^3h_j(V_j - V_{Na}) - g_Kn_j^4(V_j - V_K) - g_L(V_j - V_L) \\ +I_{FF,j} + J_{FF,j} + I_{PF,j} + I_{TF,j} + \eta_F\xi_{F,j}(t) \quad (6)$$

$$\frac{dm_j}{dt} = \alpha_{m,j}(1 - m_j) - \beta_{m,j} m_j \tag{7}$$

$$\frac{dh_j}{dt} = \alpha_{h,j}(1 - h_j) - \beta_{h,j} h_j, \tag{8}$$

$$\frac{dn_j}{dt} = \alpha_{n,j}(1 - n_j) - \beta_{n,j} n_j, \tag{9}$$

where $C = 1\ \mu F/cm^2$, $I_{F,j} = I_F\,(j = 1, 2, ..N)$ is the external stimulation current. The maximal specific conductances,the reversal potentials and the rate variables are equal to those adopted for the pyramidal cell model. In this model the onset of periodic firing occurs through an Hopf bifurcation for $I_F \cong 1.04$ $\mu A/cm^2$ with a well defined frequency ($\nu \cong 2Hz$).

The current $I_{FF,j}$ arises from the inhibitory coupling of the $j-th$ FS interneuron with the other cells, while $J_{FF,j}$ describes the current due to the electrical coupling (gap-junction) among interneurons; $I_{PF,j}$ describes the excitatory current due to the coupling with the network of pyramidal neurons, and $I_{TF,j}$ represents the excitatory current from the TC pool. These currents will be defined in the next section. To reproduce the membrane potential fluctuations each $j-th$ cell model is injected with the noisy current $\eta_F \xi_{F,j}(t)$, $\xi_{F,j}$ being an uncorrelated Gaussian random variable of zero mean and unit standard deviation $< \xi_{F,i}, \xi_{F,j} >= \delta_{ij}, i \neq j = 1, 2, 3, , N_{FS}$ and $< \xi_{P,i}, \xi_{F,j} >= 0$. The value of the η_F was chosen to get realistic amplitude of the fluctuation of membrane potential. The single compartment models of RE and TC cells were adopted from [9]. The mathematical model of the $j-th$ reticular neuron reads:

$$C\frac{dV_j}{dt} = -g_{Ca-R} m^2_{R,\infty}(V_j) h_{R,j}(V_j - V_{Ca}) - g_{LR}(V_i - V_{LR})$$
$$-g_{AHP} m_{j,AHP}(V_j - V_K) + I_{RRA,j} + I_{RRB,j} + I_{TR,j} + I_{PR,j} + \eta_R \xi_{R,j}(t) \tag{10}$$

$$\frac{dh_{R,j}}{dt} = 4.2(h_\infty(V_j) - h_{R,j})/\tau_{R,h}(V_j) \tag{11}$$

$$\frac{dm_{j,AHP}}{dt} = 0.02[Ca]_j(1 - m_{j,AHP}) - 0.025 m_{j,AHP} \tag{12}$$

$$\frac{d[Ca]_j}{dt} = -0.01 g_{Ca-R} m^2_\infty(V_j) h_{R,j}(V_j - V_{Ca}) - 0.08[Ca]_j \tag{13}$$

The maximal specific conductances and the reversal potentials are respectively: $C = 1\ \mu F/cm^2$, g_{Ca-R} =2 mS/cm^2, $V_{Ca} = 120\ mV$, $g_{LR} = 0.06\ mS/cm^2$, V_{LR} = -60 mV, $g_{AHP} = 0.3\ mS/cm^2$, V_K = -90 mV. The rate variables describing the currents are defined by: $m_{R,\infty}(V_j) = [1 + exp(-(V_j + 52)/7.4)]^{-1}$, $h_{R,\infty}(V_j) = [1 + exp((V_j + 78)/5)]^{-1}$, $\tau_{R,h}(V_j) = 100 + 500[1 + exp((V_j + 78)/3)]^{-1}$. The currents $I_{RRA,j}$, $I_{RRB,j}$ represent inhibitory coupling among RE cells, $I_{TR,j}$ and $I_{PR,j}$ describe respectively excitatory inputs from TC and pyramidal neurons. To reproduce the membrane potential fluctuations each $j-th$ cell model is injected with the noisy current $\eta_R \xi_{R,j}(t)$, $\xi_{R,j}$ being an uncorrelated Gaussian random variable of zero mean and unit standard deviation $< \xi_{R,i}, \xi_{R,j} >= \delta_{ij}, i \neq j = 1, 2, 3, , N_{RE}$. The adopted value of the parameter η_R was chosen to get realistic amplitude of the fluctuations of membrane potential.

The mathematical model of the $j-th$ thalamocortical neuron reads:

$$C\frac{dV_j}{dt} = -g_{Ca-T} m^2_{T,\infty}(V_j) h_{T,j}(V_j - V_{Ca}) - g_{LT}(V_j - V_{LT}) - g_{sag} r_j (V_j - V_{sag}) + I_{RTA,j} + I_{RTB,j} + I_{PT,j} + \eta_T \xi_{T,j}(t) \quad (14)$$

$$\frac{dh_{T,j}}{dt} = 4.2(h_{T,\infty}(V_j) - h_{T,j})/\tau_{T,h}(V_j) \quad (15)$$

$$\frac{dr_j}{dt} = (r_\infty(V_j) - r_j)/\tau_{sag}(V_j) \quad (16)$$

The maximal specific conductances and the reversal potentials are respectively: $C = 1\ \mu F/cm^2$, g_{Ca-T} =2.5 mS/cm^2, $V_{Ca} = 120\ mV$, $g_{LT} = 0.025\ mS/cm^2$, V_{LR} = -75 mV, $g_{sag} = 0.04\ mS/cm^2$, V_{sag} = -40 mV. The rate variables describing the currents are defined by: $m_{T,\infty}(V_j) = [1 + exp(-(V_j + 59)/6.2)]^{-1}$, $h_{T,\infty}(V_j) = [1+exp((V_j+81)/4.4)]^{-1}$, $\tau_h(V_j) = 30+220[1+exp((V_j+78)/3)]^{-1}$, $r_\infty(V_j) = [1+exp((V_j+75)/5.5)]^{-1}$, $\tau_{sag}(V_j) = 20+1000[exp((V_j+71.5)/14.2)+ exp(-(V_j + 89)/11.6)]^{-1}$. The currents $I_{RTA,j}$, $I_{RTB,j}$ represent inhibitory inputs due to the coupling with the network of RE neurons , $I_{PT,j}$ describes excitatory inputs from pyramidal neurons to TC cells. To reproduce the membrane potential fluctuations each $j-th$ cell model is injected with the noisy current $\eta_T \xi_{T,j}(t)$, $\xi_{T,j}$ being an uncorrelated Gaussian random variable of zero mean and unit standard deviation $< \xi_{T,i}, \xi_{T,j} >= \delta_{ij}, i \neq j = 1, 2, 3, , N_{TC}$. The adopted value of the parameter η_T was chosen to get realistic amplitude of the fluctuations of membrane potential.

The reason of using a single compartment model of each cell is motivated by computational constraints. The simulation will be performed by using up to 180 coupled neuron models, and this requires a high computational cost. Therefore, for the aim of the present work, the choice of using a single compartment biophysical model of each cell is a good compromise between two requirements: computational advantages and realistic network of coupled neurons.

2.2 Synaptic Coupling

The excitatory synaptic coupling among pyramidal cells is assumed to be all-to-all. The excitatory synaptic current acting on the $j-th$ pyramidal cell is defined by

$$I_{PP,j} = -\frac{1}{N_{PY}-1}\sum_{k\neq j} g_e s_{PP,k}(t)(V_j - V_{PP}) \tag{17}$$

where $g_e = 0.5mS/cm^2$ represents the maximal amplitude of the excitatory coupling, the function $s_{PP,k}(t)$ describes the time evolution of the postsynaptic current and V_{PP} is the corresponding reversal potential. According to [8] the time evolution of $s_{PP,k}(t)$ is described by

$$\frac{ds_{PP,k}(t)}{dt} = T(V_k)(1 - s_{PP,k}) - s_{PP,k}/\tau_e \tag{18}$$

where $T(V_k) = 5(1 + tanh(V_k/4)$ and $\tau_e = 2ms$ is the decay time constant.

Similarly the inhibitory synaptic coupling among FS interneurons is assumed to be all-to-all and the synaptic current on the $j-th$ interneuron reads

$$I_{FF,j} = -\frac{1}{N_{FS}-1}\sum_{k\neq j} g_i s_{FF,k}(t)(V_j - V_{FF}) \tag{19}$$

where $g_i = 0.25mS/cm^2$ represents the maximal amplitude of the inhibitory coupling and V_{FF} is the corresponding reversal potential. The time evolution of $s_{FF,k}(t)$ is described by

$$\frac{ds_{FF,k}(t)}{dt} = T(V_k)(1 - s_{FF,k}) - s_{FF,k}/\tau_i \tag{20}$$

where $T(V_k) = 2(1 + tanh(V_k/4)$ and $\tau_1 = 10ms$ is the decay time constant.

The pyramidal cells excite the network of FS cells and the corresponding excitatory current acting on the $j-th$ interneuron is defined as

$$I_{PF,j} = -\frac{1}{N_{PY}-1}\sum_{k\neq j} g_{PF} s_{PF,k}(t)(V_j - V_{PF}) \tag{21}$$

where g_{PF} represents the maximal amplitude of the excitatory coupling and $V_{PF} = V_{PP}$ is the corresponding reversal potential. The time evolution of $s_{PF,k}(t)$ is driven by

$$\frac{ds_{PF,k}(t)}{dt} = T(V_k)(1 - s_{PF,k}) - s_{PF,k}/\tau_e \tag{22}$$

where $T(V_k) = 5(1 + tanh(V_k/4)$ and $\tau_e = 2ms$ is the decay time constant.

The network of FS interneurons feedback inhibition to the pyramidal neurons and the inhibitory current of the $j-th$ cell is given by

$$I_{FP,j} = -\frac{1}{N_{FS}-1}\sum_{k\neq j} g_{FP} s_{FP,k}(t)(V_j - V_{FP}) \tag{23}$$

where g_{FP} represents the maximal amplitude of the inhibitory coupling and $V_{FP} = V_{FF}$ is the corresponding reversal potential. The time evolution of $s_{FP,k}(t)$ is determined by

$$\frac{ds_{FP,k}(t)}{dt} = T(V_k)(1 - s_{FP,k}) - s_{FP,k}/\tau_i \tag{24}$$

where $T(V_k) = 2(1 + tanh(V_k/4)$ and $\tau_i = 10ms$ is the decay time constant.

The electrical coupling among FS interneurons is all-to-all and the corresponding current on the $j - th$ cell is defined as

$$J_{FF,j} = \frac{1}{N_{FS} - 1} \sum_{k \neq j} g_{el}(V_j - V_k) \tag{25}$$

where g_{el} is the coupling amplitude. The parameters values g_i, g_e, g_{PF}, g_{FP} are those adopted in [8] .

The excitatory coupling due to TC network on FS neurons is described by an AMPA current

$$I_{TF,j} = -\frac{1}{N_{TC}} \sum_k g_{AMPA(T-F)} s_{TF,k}(t)(V_j - V_{AMPA}) \tag{26}$$

where $g_{AMPA(T-F)}$ represents the maximal amplitude of the excitatory coupling and $V_{AMPA} = 0\ mV$ is the corresponding reversal potential.The time evolution of the synaptic variable $s_{TF,k}(t)$ is described by

$$\frac{ds_{TF,k}(t)}{dt} = s_\infty(V_k)(1 - s_{TF,k}) - s_{TF,k}/\tau_T \tag{27}$$

where $s_\infty(V_k) = 2[1 + exp(-(V_k + 45)/2)]^{-1}$, $\tau_T = 10\ ms$.

The excitatory coupling due to TC network on PY neurons is described by an AMPA current

$$I_{TP,j} = -\frac{1}{N_{TC}} \sum_k g_{AMPA(T-P)} s_{TP,k}(t)(V_j - V_{AMPA}) \tag{28}$$

where $g_{AMPA(T-P)}$ represents the maximal amplitude of the excitatory coupling and $V_{AMPA} = 0\ mV$ is the corresponding reversal potential.The time evolution of the synaptic variable $s_{TP,k}(t)$ is described by

$$\frac{ds_{TP,k}(t)}{dt} = s_\infty(V_k)(1 - s_{TP,k}) - s_{TP,k}/\tau_T \tag{29}$$

where $s_\infty(V_k) = 2[1 + exp(-(V_k + 45)/2)]^{-1}$, $\tau_T = 10\ ms$.

The inhibitory coupling among RE neurons are characterized by GABA-A and GABA-B synapses, defined by the following equations

$$I_{RRA,j} = -\frac{1}{N_{RE}} \sum_k g_{GABA-A} s_{RA,k}(t)(V_j - V_{GABA-A}) \tag{30}$$

where $g_{GABA-A} = 0.5\ mS/cm^2$ represents the maximal amplitude of the inhibitory coupling and V_{GABA-A} = -75 mV is the corresponding reversal potential.The time evolution of the synaptic variable $s_{RA,k}(t)$ is described by

$$\frac{ds_{RA,k}(t)}{dt} = x_\infty(V_k)(1 - s_{RA,k}) - s_{RA,k}/\tau_A \tag{31}$$

where $x_\infty(V_k) = 2[1 + exp(-(V_j + 45)/2)]^{-1}$ and $\tau_A = 12.5\ ms$

$$I_{RRB,j} = -\frac{1}{N_{RE}} \sum_k g_{GABA-B} s_{RB,k}(t)(V_j - V_{GABA-B}) \tag{32}$$

where $g_{GABA-B} = 0.1\ mS/cm^2$ represents the maximal amplitude of the inhibitory coupling and V_{GABA-B} = -90 mV is the corresponding reversal potential.The time evolution of the synaptic variable $s_{RB,k}(t)$ is described by

$$\frac{ds_{RB,k}(t)}{dt} = s_\infty(x_{RB,k})(1 - s_{RB,k}) - s_{RB,k}/\tau_B \tag{33}$$

$$\frac{dx_{RB,k}(t)}{dt} = x_\infty(V_k)(1 - x_{RB,k}) - x_{RB,k}/\tau_{x,B} \tag{34}$$

where $s_\infty(x_{RB,k}) = 0.01[1+exp(-(x_{RB,k}-1/e)/0.02)]^{-1}$, $\tau_B = 200\,ms\ x_\infty(V_k) = 5[1 + exp(-(V_j + 45)/2)]^{-1}$ and $\tau_{x,B} = 100\ ms$,

The excitatory coupling due to TC network on RE neurons is described by an AMPA current

$$I_{TR,j} = -\frac{1}{N_{TC}} \sum_k g_{AMPA(T-R)} s_{TR,k}(t)(V_j - V_{AMPA}) \tag{35}$$

where $g_{AMPA(T-R)} = 0.02\ mS/cm^2$ represents the maximal amplitude of the excitatory coupling and $V_{AMPA} = 0\ mV$ is the corresponding reversal potential.The time evolution of the synaptic variable $s_{TR,k}(t)$ is described by

$$\frac{ds_{TR,k}(t)}{dt} = s_\infty(V_k)(1 - s_{TR,k}) - s_{TR,k}/\tau_T \tag{36}$$

where $s_\infty(V_k) = 2[1 + exp(-(V_k + 45)/2)]^{-1}$, $\tau_T = 10\ ms$

The excitatory coupling due to PY network on RE neurons is described by an AMPA current

$$I_{PR,j} = -\frac{1}{N_{PY}} \sum_k g_{AMPA(P-R)} s_{PR,k}(t)(V_j - V_{AMPA}) \tag{37}$$

where $g_{AMPA(P-R)}$ represents the maximal amplitude of the excitatory coupling and $V_{AMPA} = 0\ mV$ is the corresponding reversal potential.The time evolution of the synaptic variable $s_{PR,k}(t)$ is described by

$$\frac{ds_{PR,k}(t)}{dt} = T(V_k)(1 - s_{PR,k}) - s_{PR,k}/\tau_e \tag{38}$$

where $T(V_k) = 5(1 + tanh(V_k/4)$ and $\tau_e = 2ms$ is the decay time constant.

The inhibitory coupling due to RE neurons on TC neuron are characterized by GABA-A and GABA-B synapses, defined by the following equations

$$I_{RTA,j} = -\frac{1}{N_{RE}} \sum_k g^{RT}_{GABA-A} s_{RA,k}(t)(V_j - V_{GABA-A}) \tag{39}$$

where g^{RT}_{GABA-A} represents the maximal amplitude of the inhibitory coupling and V_{GABA-A} = -75 mV is the corresponding reversal potential.The time evolution of the synaptic variable $s_{RA,k}(t)$ is described by

$$\frac{ds_{RA,k}(t)}{dt} = x_\infty(V_k)(1 - s_{RA,k}) - s_{RA,k}/\tau_A \tag{40}$$

where $x_\infty(V_k) = 2[1 + exp(-(V_j + 45)/2)]^{-1}$ and $\tau_A = 12.5\ ms$

$$I_{RTB,j} = -\frac{1}{N_{RE}} \sum_k g^{RT}_{GABA-B} s_{RB,k}(t)(V_j - V_{GABA-B}) \tag{41}$$

where g^{RT}_{GABA-B} represents the maximal amplitude of the inhibitory coupling and V_{GABA-B} = -90 mV is the corresponding reversal potential.The time evolution of the synaptic variable $s_{RB,k}(t)$ is described by

$$\frac{ds_{RB,k}(t)}{dt} = s_\infty(x_{RB,k})(1 - s_{RB,k}) - s_{RB,k}/\tau_B \tag{42}$$

$$\frac{dx_{RB,k}(t)}{dt} = x_\infty(V_k)(1 - x_{RB,k}) - x_{RB,k}/\tau_{x,B} \tag{43}$$

where $s_\infty(x_{RB,k}) = 0.01[1+exp(-(x_{RB,k}-1/e)/0.02)]^{-1}$, $\tau_B = 200\,ms\ x_\infty(V_k) = 5[1 + exp(-(V_j + 45)/2)]^{-1}$ and $\tau_{x,B} = 100\ ms$,

The excitatory coupling due to PY network on TC neurons is described by an AMPA current

$$I_{PT,j} = -\frac{1}{N_{PY}} \sum_k g_{AMPA(P-T)} s_{PT,k}(t)(V_j - V_{AMPA}) \tag{44}$$

where $g_{AMPA(P-T)}$ represents the maximal amplitude of the excitatory coupling and $V_{AMPA} = 0\ mV$ is the corresponding reversal potential.The time evolution of the synaptic variable $s_{PR,k}(t)$ is described by

$$\frac{ds_{PT,k}(t)}{dt} = T(V_k)(1 - s_{PT,k}) - s_{PT,k}/\tau_e \tag{45}$$

where $T(V_k) = 5(1 + tanh(V_k/4)$ and $\tau_e = 2ms$ is the decay time constant. The parameter values describing the amplitude of the synaptic current among neurons of the thalamus were those reported in [9]. The remaining values describing the synaptic currents from cortical and thalamic cells were chosen to get realistic amplitude of the postsynaptic potentials.

3 Results

Let us start first by investigating how the the whole cortico-thalamo-cortical network behave in absence of coupling between cortical and thalamic compartments. The corresponding results are reported in the left panel of figure 2.

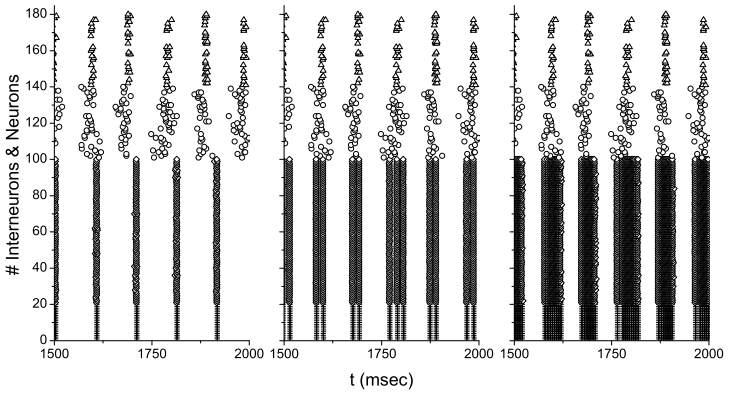


Fig. 2. Effects of coupling betweeen thalamic neuron and pyramidal neurons on the networks dynamics. Left panel: $g_{AMPA(T-P)} = 0$, $g_{AMPA(T-F)} = 0$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$,. Middle panel: $g_{AMPA(T-P)} = 0.2mS/cm^2$, $g_{AMPA(T-F)} = 0$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$. Right panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 0$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$. For all panels it is: $g_e = 0.5mS/cm^2$, $g_i = 0.25mS/cm^2$, $g_{el} = 0$, $I_P = 3.5\mu A/cm^2$, $I_F = 0.5\mu A/cm^2$, $g^{RT}_{GABA-A} = 0.15mS/cm^2$, $g^{RT}_{GABA-B} = 0.05mS/cm^2$, $N_{PY} = 80$, $N_{FS} = 20$, $N_{TC} = 40$, $N_{RE} = 40$. For all panels the plus symbols represent the FS interneurons, the diamonds represent the pyramidal cells, the open circles represent the TC neurons and the open triangles the RE neurons.

FS and pyramidal neurons fire in close synchrony and a qualitatively similar behaviour is exhibited by RE and TC cells (in this last case the level of network synchrony is smaller than that for the cortical compartment). In the middle panel are reported the results obtained in the case in which the pyramidal cell receive the excitatory synaptic inputs from TC cells ($g_{AMPA(T-P)} = 0.2mS/cm^2$). In this case there is an increment of the spiking activity of pyramidal neurons. The increase of the amplitude of the synaptic input from TC cells to pyramidal neurons promotes the generation of seizure-like dynamics (right panel of figure 2). This behaviour is in agreement with the experimental results described in [5]. An important point concerns the impact of the FS cells on the network dynamics when the synaptic input from TC to FS cells is set on. The results for the case $g_{AMPA(T-F)} = 0.5mS/cm^2$ are reported in the left panel of figure 3. Comparing these results with those reported in the right panel of figure 2 it follows that the main effect of this additional coupling is to enlarge the time window where the bursting of FS cells occur.

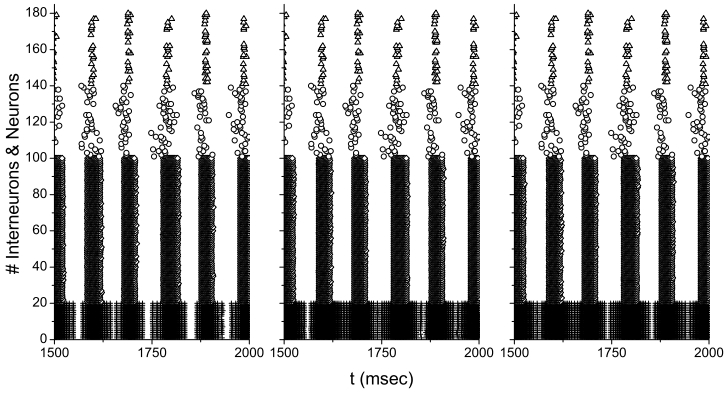


Fig. 3. Effects of coupling betweeen thalamic neuron and pyramidal neurons on the networks dynamics. Left panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 0.5mS/cm^2$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$, $g_{el} = 0$. Middle panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 1.2mS/cm^2$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$, $g_{el} = 0$. Right panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 1.2mS/cm^2$, $g_{AMPA(P-T)} = 0$, $g_{AMPA(P-R)} = 0$, $g_{el} = 0.15mS/cm^2$. For all panels it is: $g_e = 0.5mS/cm^2$, $g_i = 0.25mS/cm^2$, $I_P = 3.5\mu A/cm^2$,$g^{RT}_{GABA-A} = 0.15mS/cm^2$, $g^{RT}_{GABA-B} = 0.05mS/cm^2$, $I_F = 0.5\mu A/cm^2$, $N_{PY} = 80$, $N_{FS} = 20$, $N_{TC} = 40$, $N_{RE} = 40$. For all panels the plus symbols represent the FS interneurons, the diamonds represent the pyramidal cells, the open circles represent the TC neurons and the open triangles the RE neurons.

However, as expected, the increase of the inhibitory inputs determines a reduction of the time window where the bursting of the pyramidal cells occurs. Increasing more the coupling amplitude ($g_{AMPA(T-F)} = 1.2mS/cm^2$) promotes the increase of the firing activity of the FS cells (see middle panel of figure 3), but the activity of the pyramidal cell is practically unaffected. Similar results were found when the electrical coupling among FS cells was set on (see right panel of figure 3). Let us now study how the presence of the excitatory synaptic inputs from pyramidal cells to RE and TC neurons affects the network dynamics. The corresponding data reported in figure 4 show that both RE and TC cells now fire more synchronoulsly; moreover a reduction of the firing activity of the FS cells also occur (see for comparison the data in the middle panel of figure 3). Also in this situation the pyramidal cells exhibit seizure like behaviour. The increase of the amplitude of the synaptic inputs between RE and TC cells has a dramatic effect on the network dynamics. The corresponding results are reported in the right panel of figure 4 and clearly show that the seizure like behavior is terminated.

In addition the data show a complete cessation of spiking activity of the thalamic cells. Why this occurs? To respond to this question it is important to know how RE and TC cell models behave when receive excitatory (or inhibitory inputs). To this aim let us consider a single RE (or TC) cell injected with a

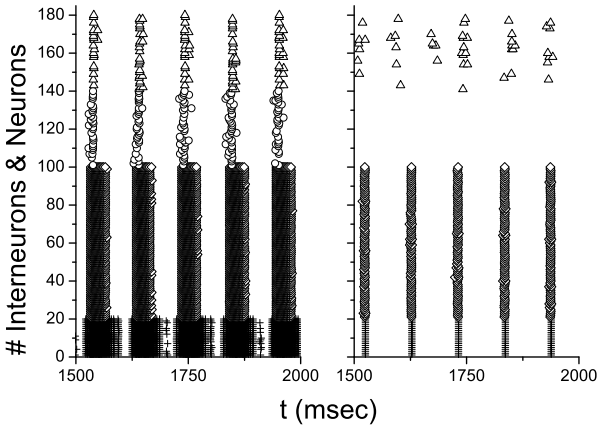


Fig. 4. Effects of coupling betweeen thalamic neuron and pyramidal neurons on the networks dynamics. Left panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 0.5mS/cm^2$, $g_{AMPA(P-T)} = 0.05mS/cm^2$, $g_{AMPA(P-R)} = 0.02mS/cm^2$, $g^{RT}_{GABA-A} = 0.15mS/cm^2$, $g^{RT}_{GABA-B} = 0.05mS/cm^2$. Right panel: $g_{AMPA(T-P)} = 0.8mS/cm^2$, $g_{AMPA(T-F)} = 1.2mS/cm^2$, $g_{AMPA(P-T)} = 0.05mS/cm^2$, $g_{AMPA(P-R)} = 0.02mS/cm^2$, $g^{RT}_{GABA-A} = 1.1mS/cm^2$, $g^{RT}_{GABA-B} = 0.9mS/cm^2$. For all panels it is: $g_e = 0.5mS/cm^2$, $g_i = 0.25mS/cm^2$, $g_{el} = 0$, $I_P = 3.5\mu A/cm^2$, $I_F = 0.5\mu A/cm^2$, $N_{PY} = 80$, $N_{FS} = 20$, $N_{TC} = 40$, $N_{RE} = 40$. For all panels the plus symbols represent the FS interneurons, the diamonds represent the pyramidal cells, the open circles represent the TC neurons and the open triangles the RE neurons.

depolarizing (hyperpolarizing) current. Let us first consider the RE cell that, for the adopted parameter values, generates action potential spontaneously at a frequency of about 8 Hz. When this cell is injected with a depolarizing current the amplitude of the action potential decreses as the amplitude current grows (the corresponding frequency exhibit small changes). When the current amplitude is greater than 1.1 mA/cm^2, the firing disappears. Therefore, for the RE cell model a depolarizing input depress the firing activity. If the RE cell model is injected with an hyperpolarizing current, the amplitude of the action potential increases. However, for amplitudes of the injected current smaller than -1 mA/cm^2 the firing ceases (data not shown). Let us now consider the TC cell model. In this case, for the adopted parameter value, the value of the membrane potential of a single TC cell is at resting (about $-55mV$). If the cell is injected with a depolarizing current the values of the membrane potential increases but no firing occurs. If the TC cell is injected with an hyperpolarizing current of amplitude -0.7 mA/cm^2 then a firing activity starts (with a frequency of about $2.5Hz$). As the current gets smaller values both the amplitude of the action potential and the firing frequency decrease; the firing disappears for current amplitude smaller than -1.4 mA/cm^2 (data not shown). The single pyramidal neuron (or the FS interneuron) model when injected with a constant current behaves more regularly than RE or TC cells [8,10]. Then, it can be shown qualitatively that the seizure

termination (see figure 4) is a direct consequence of the response properties of the single RE and TC neuron models to depolarizing (hyperpolarizing) inputs discussed before.

Taken together these results are qualitatively in keeping with those described in [5]. In particular in this experimental work it was shown that the inhibition of thalamocortical neurons interrupted seizures, and this indicates that a suitable modulation of the activity of the cortico-thalamo-cortical network could be used to control seizures generation and termination.

4 Conclusions

The main goal of this paper was the understanding of some possible mechanisms controlling epileptic seizures dynamics. Recently, experimental results pointed out that thalamic inputs modulate seizure dynamics: i.e. they can promote generation and termination of ictal activity [5]. Motivated by these experimental data, we studied the dynamical behaviour of a biophysical inspired network of four coupled populations of cells: the first population is composed by coupled FS interneurons (coupled by inhibitory and electrical synapses), the second one is constituted by coupled pyramidal cells (coupled by excitatory synapses), the third is composed by thalamic reticular neurons RE (coupled by inhibitory synapses), while the last pool is composed by thalamocortical cells TC. In particular, our attention was focused to study how alteration of the coupling among thalamic and cortical neurons affects the whole network firing activity. The numerical simulations have shown that the increase of the amplitude of the excitatory coupling from TC neurons to pyramidal cells promotes the generation of seizure-like behaviour (see figure 2). Furthermore, the addition of the excitatory coupling from TC neurons to FS interneurons enlarges the time window where the bursting of FS cells occurs. As a consequence, the fast bursting regime of the pyramidal cells is reduced. The presence of excitatory synaptic inputs from pyramidal neurons to RE and TC neurons produce a synchronization of the firing activity of both RE and TC cells. A remarkable result is obtained when the inhibitory coupling on TC cells from the RE pool is increased. Indeed the spiking activity of the TC neurons is absent and we observed a complete cessation of the bursting regime of pyramidal neurons (see figure 4). In conclusion this computational study have clearly shown that the cortico-thalamo-cortical network is capable of promoting (or inhibiting) cortical ictal activity. These findings are in agreement with the experimental results described in [5].

References

1. Zhang, Z.J., Valiante, T.A., Carlen, P.L.: Transition to seizure: From “macro”- to “micro”-mysteries. Epilepsy Research 97, 290–299 (2011)
2. Lado, F.A., Moshe, F.L.: How do seizures stop? Epilepsia 49, 1651–1664 (2008)
3. Truccolo, W., et al.: Single-neuron dynamics in human focal epilepsy. Nature Neurosci. 14, 635–643 (2011)

4. Huguenard, J.R., Prince, D.A.: Basic mechanisms of epileptic discharges in the thalamus. In: Steriade, M., Jones, E.G., McCormick, D. (eds.) The Thalamus: Experimental and Clinical Aspects, pp. 295–330. Elsevier (1997)
5. Paz, J.T., Davidson, T.J., Frechette, E.S., Delord, B., Parada, I., Peng, K., Deisseroth, K., Huguenard, J.R.: Closed-loop optogenetic control of thalamus as a tool for interrupting seizures after cortical injury. Nature Neuroscience 16, 64–70 (2013)
6. Amritkar, R.E., Rangarajan, G.: Spatially synchronous extinction of species under external forcing. Phys. Rev. Lett. 96, 258102 (2006)
7. Krishnan, G.P., Bazhenov, M.: Ionic dynamics mediate spontaneous termination of seizures and postictal depression state. J. Neurosci. 31(24), 8870–8882 (2011)
8. Olufsen, M., Whittington, M., Camperi, M., Kopell, N.: New roles for the gamma rhythm: population tuning and preprocessing for the beta rhythm. J. Comp. Neurosci. 14, 33–54 (2003)
9. Golomb, D., Wang, X.-J., Rinzel, J.: Synchronization properties of spindle oscillations in a thalamic reticular nucleus model. J. Neurophysiol. 72, 1109–1126 (1994)
10. Chillemi, S., Barbi, M., Di Garbo, A.: A network of pyramidal neurons is sensitive to the timing of its excitatory inputs. Neurocomputing 74, 1159–1164 (2011)

Modelling Prior and Retrospective Awareness of Actions

Dilhan J. Thilakarathne and Jan Treur

VU University Amsterdam, Agent Systems Research Group
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{d.j.thilakarathne,j.treur}@vu.nl
http://www.cs.vu.nl/~treur

Abstract. Agents often may prepare for and perform actions without being conscious of these processes. However, in other cases, at some point in time the agent can develop some awareness state relating to the action. This can be an awareness state prior to the execution of the action. An awareness state can also develop in retrospect, after the action was performed. In this paper a neurologically inspired agent model is introduced that is able to make such distinctions. Scenarios are covered in which actions are prepared without being conscious at any point in time. Also scenarios are covered in which the agent develops proir awareness or retrospective awareness, or both. When prior awareness is developed it may be the case that this awareness has a decisive effect on actually executing the action, but it may equally well be the case that the awareness state has no effect on whether the action is performed. All these variations have been illustrated by a wide variety of simulation experiments.

Keywords: awareness, ownership, prior, retrospective.

1 Introduction

Agents often may prepare for and perform actions without being conscious of these preparation and execution processes. However, in other cases, at some point in time the agent can develop some awareness state relating to the action. This can be an awareness state prior to the execution of the action, or an awareness state in retrospect, after the action was performed.

When a prior awareness state occurs, the agent becomes aware of going to perform the action. Whether or not the agent's decision to actually perform the action depends on this prior awareness state has been debated extensively in the literature; e.g., [1–4]. Having such a prior awareness state still may leave open whether the agent is able to consciously decide to perform or not to perform the action. For example, is still some form of veto-ing of the action possible? In principle, the awareness state may play the role of generating a kind of green light for execution of the action. However, equally well the prior awareness state may just play the role of a warning for the agent to be prepared that the action will happen (anyway). It has been found that for certain types of actions the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 62–73, 2013.

decision to perform it is already made at least hundreds of milliseconds (and even up to 10 seconds) before any awareness state occurs; e.g. [1, 2]. These findings may suggest that prior awareness often will have no effect on the decision. But this may strongly depend on the type of action. For example, it will be difficult to believe that the action of buying a car or a house remain unconscious and may not be amendable to veto-ing based on awareness states.

An awareness state can also develop in retrospect, after the action was performed ('what have I done?'). Such a retrospective awareness state often relates to acknowledging others from and taking responsibility for having performed the action. It may also play an important role in learning: by evaluating the obtained effect in a conscious manner in order to improve a next performance of the action.

In this paper a neurologically inspired agent model is introduced that is able to make such distinctions. Scenarios are covered in which actions are prepared without being conscious at any point in time. Also scenarios are covered in which the agent develops prior awareness or retrospective awareness, or both. When prior awareness is developed it may be the case that this awareness has a decisive effect on actually executing the action, but it may equally well be the case that the awareness state has no effect on whether the action is performed. All these variations have been illustrated by a wide variety of simulation experiments. The paper is organised as follows. In Section 2, some background from neurological literature is presented. Next, in Section 3 the agent model is introduced. Some of the simulation experiments are discussed in Section 4. Section 5 concludes the paper with a discussion.

2 Background

In the cognitive and neurological literature the notions of awareness and ownership of an action have received much attention. For example, in how far does a person attribute an action to him or herself, or to another person. Persons suffering from schizophrenia may easily attribute self-generated actions to (real or imaginary) other persons. One of the issues that have turned out to play an important role both in the execution decisions for an action, and in its attribution, is the prediction of the (expected) effects of the action, based on internal simulation starting from the preparation of the action (e.g., [5, 6]). If these predicted effects are satisfactory, this may entail a 'go' decision for the execution of the action, thus exerting control over action execution. In contrast, less satisfactory predicted effects may lead to a 'no go' decision. Predicted action effects also play an important role in attribution of the action to an agent after it has been performed. In neurological research it has been found that poor predictive capabilities are a basis for false attributions of actions, for example, for patients suffering from schizophrenia; (e.g., [7, 8]). In recent literature it has been reported that the predicted sensory effect and the sensed actual effect are integrated with each other as a basis for proper attribution of the action (e.g., [7–9]).

Another element, put forward in [9], is the distinction between action awareness based on prediction (prior to execution), and action awareness based on inference after execution of the action (in retrospect):

> 'Our results suggest that both predictive and inferential processes contribute to the conscious awareness of operant action. The relative contribution of each of these processes seems to be context dependent. When we can predict the consequences of our actions, as in a high action-effect contingency block, the awareness of action reflects these predictions. This would provide us with a predictive sense of our own agency. In addition, our results show clear evidence that inferential processes also influence the conscious awareness of operant action. (...) The interaction between predictive and inferential processes is of particular interest. (...) The time course over which information about action is built up may be an important clue to this interaction. (...) Sensory feedback provides more precise evidence about actions and their effects. This evidence becomes available only after a short sensory delay, but can then be transferred to memory. Thus, reliable and enduring sensory evidence replaces short-lived predictive estimates. We suggest that awareness of action therefore switches from a predictive to an inferential source as the action itself occurs, and as sensory information becomes available.' ([9], pp. 142-143)

Here Moore and Haggard point at the timing aspect of awareness states of actions. In the agent model introduced below this has been addressed by distinguishing prior and retrospective awareness states.

3 Description of the Agent Model

Having neurological evidence on awareness (together with debates concerning its illusionary character in the literature) this section presents an agent model that will be used in agent driven applications where the awareness is paramount (or necessary) in terms of decision making and justifications of actions through communication. More specifically, in problem domains concerning performing or learning specific healthy behaviours or lifestyles having an idea about the extent of the awareness about decisions is an interesting and important issue; this model (with further refinements and customization where needed) may provide the fundamentals. Further this model may be useful in medical domains where through simulations students can learn (or teach) and compare the phenomenal effects of certain scenarios. An overview of the postulated cognitive agent model is presented in the Figure 1 below. The followings have been incorporated in the process of its modeling:

(1) action effect-prediction (through as-if body loop): sensory representation of effect b_i is affected from preparation of an action a_i
(2) preparation for action a_i is affected by sensory representation of s, prior-awareness, and feeling of effect prediction of action a_i
(3) a prior ownership state depends on preparation for action a_i, predicted effects b_i of a_i, stimulus s, retrospective ownership and context c
(4) a prior-awareness state depends on sensory representation of stimulus s, feeling of predicted effect, prior ownership and retrospective awareness

(5) execution of action a_i is affected by effects from prior-awareness, prior ownership and preparation for action a_i
(6) a prior ownership state and prior awareness state exert control over the execution of a prepared action (go/no-go decision, vetoing)
(7) suppressing the sensory representation of effect b_i by the both prior-ownership and retrospective-awareness after the action a_i was initiated
(8) suppressing the prior ownership state when the retrospective ownership state is getting developed
(9) suppressing the prior-awareness state when the retrospective awareness state is getting developed
(10) a retrospective ownership state depends on co-occurrence of predicted action effects and action effects sensed afterwards
(11) a retrospective-awareness state depends on action effects sensed by execution of action a_i, retrospective ownership, and prior-awareness
(12) a retrospective ownership state and retrospective-awareness are internal states that also can lead to acknowledging authorship of the action (individually), for example in social context
(13) execution of an action a_i affect the stimulus s in the world

The state labels used in the model are summarized in Table 1. The model adopts parts of the model presented in [10] but extends this by introducing the prior and retrospective awareness states for actions. These awareness states are taken specific for a given action a, effect b, context c, and stimulus s (triggering preparation of a). When the context c is $self$, an awareness state for c indicates self-attribution awareness, whereas for context c an observed agent B, it indicates awareness of attribution of the action to B.

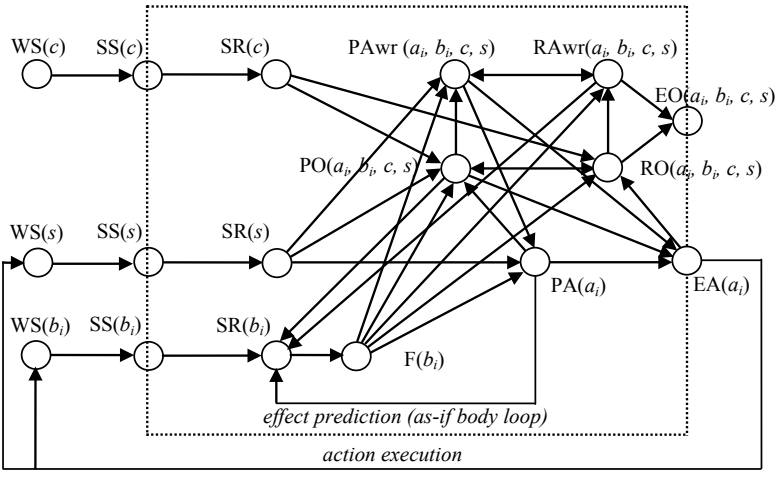


Fig. 1. Overview of the cognitive agent model

In this model when a detectable stimulus is occurring, the agent may prepare for more than one action internally. The model is having the innate ability to mentally explore multiple action options in response to a given stimulus before selecting an appropriate action to execute it. In these processes awareness may or may not play a role. Different options are labeled with indices: a_x and b_x (see also Figure 1).

Table 1. State labels used

notation	description
WS(*W*)	world state *W* (*W* is a context *c*, stimulus *s*, or effect *b*)
SS(*W*)	sensor state for *W*
SR(*W*)	sensory representation of *W*
PA(*a*)	preparation for action *a*
EA(*a*)	execution of action *a*
F(*b*)	feeling of action *a* after: as-if loop or action execution
PO(*a, b, c, s*)	prior ownership state for action *a* with *b*, *c*, and *s*
RO(*a, b, c, s*)	retrospective ownership state for *a* with *b*, *c*, and *s*
EO(*a, b, c, s*)	communication of ownership of *a* with *b*, *c*, and *s*
PAwr(*a, b, c, s*)	prior-awareness state for action *a* with *b*, *c*, and *s*
RAwr(*a, b, c, s*)	retrospective-awareness state for action *a* with *b*, *c*, and *s*

Action prediction, expressed in (1) above, is modeled by the connection from the PA(a_i) to the SR(b_i) [11, 12]. In accordance with (2) above, PA(a_i) is affected by SR(b_i), F(b_i) via as-if body loop or body loop, and the PAwr(a_i, b_i, c, s). From (3) above, it expresses that a PO(a_i, b_i, c, s) is affected by the PA(a_i), F(b_i), SR(s), SR(c), and RO(a_i, b_i, c, s). Similar to the above explanations: (4), (5), (10), and (11) are self explainable. The control exerted by the PO(a_i, b_i, c, s), and PAwr(a_i, b_i, c, s) expressed in (6) above, is modeled by the connection from PO(a_i, b_i, c, s) and PAwr(a_i, b_i, c, s) to EA(a_i). Furthermore; due to the link from PAwr(a_i, b_i, c, s) to EA(a_i) it is facilitating a moderating role in some circumstances preventing certain actions initiated by unconscious compounds from being executed [13, 14]. In accordance with (7), SR(b_i) will be getting suppressed by both PO(a_i, b_i, c, s) [15, 16] and RAwr(a_i, b_i, c, s) [17, 18] after the EA(a_i) was initiated. Further, in line with (8) above, PO(a_i, b_i, c, s) gets suppressed by RO(a_i, b_i, c, s) once RO(a_i, b_i, c, s) got strengthened by integrating the predicted sensory effect and the sensed actual effect. The state PAwr(a_i, b_i, c, s) will be suppressed by RAwr(a_i, b_i, c, s) expressed as in (9). Finally, acknowledging of ownership, expressed in (12) above, is modeled by the connection from the RO(a_i, b_i, c, s) and RAwr(a_i, b_i, c, s) to the EO(a_i, b_i, c, s).

Connections between state properties (the arrows in Figure 1) have weights ω_k, as indicated in Table 2. In this table the column LP refers to the (temporally) Local Properties LP1 to LP13. A weight ω_k has a value between -1 and 1 and may depend on the specific context c, stimulus s, action a and/or effect state b involved. By varying these connection strengths, different possibilities for the repertoire offered by the model can be realised. Note that usually weights are

Table 2. Overview of the connections and their weights

from states	to state	weights	LP
SS(W)	SR(W)	ω_1	LP1
PA(a), PO(a,b,c,s), SS(b), RAwr(a,b,c,s)	SR(b)	ω_2, ω_3, ω_4, $\omega_{\cdot}$	LP2
SR(s), F(b), PAwr(a,b,c,s)	PA(a)	ω_5, ω_6, ω_7	LP3
SR(c), SR(s), F(b), PA(a), RO(a,b,c,s)	PO(a,b,c,s)	ω_8, ω_9, ω_{10}, ω_{11}, ω_{20o}	LP4
SR(b)	F(b)	ω_{12}	LP5
PO(a,b,c,s), PA(a), PAwr(a,b,c,s)	EA(a)	ω_{13}, ω_{14}, ω_{15}	LP6
EA(a)	WS(b)	ω_{16}	LP7
WS(W)	SS(W)	ω_{17}	LP8
SR(c), F(b), PO(a,b,c,s), EA(a)	RO(a,b,c,s)	ω_{18}, ω_{19}, ω_{20}, ω_{21}	LP9
RO(a,b,c,s), RAwr(a,b,c,s)	EO(a,b,c,s)	ω_{22}, ω_{23}	LP10
SR(s), F(b), PO(a,b,c,s), RAwr(a,b,c,s)	PAwr(a,b,c,s)	ω_{24}, ω_{25}, ω_{26}, ω_{27o}	LP11
PAwr(a,b,c,s), F(b), RO(a,b,c,s)	RAwr(a,b,c,s)	ω_{27}, ω_{28}, ω_{29}	LP12
EA(a)	WS(s)	ω''	LP13

assumed non-negative, except for the inhibiting connections, such as ω_3, ω', ω_{20o}, ω_{27o}, and ω''.

The dynamics following the connections between the states in Figure 1 have been designed. Table 3 summarizes the LP1 to LP13 in format of LEADSTO [19]. The time delay defined in LEADSTO is taken as a uniform time step Δt here. Parameter γ is a speed factor, indicating the speed by which an activation level is updated upon received input from other states. During processing, each state property has a strength represented by a real number between 0 and 1; variables V (possibly with subscripts) run over these values.

4 Simulation Results

In this section simulation experiments for a number of example scenarios are discussed. First a scenario is addressed where the prepared action has satisfactory predicted effects and therefore is executed. Next, a case is considered where the prepared action lacks positive predicted effects, and is therefore not executed. The third case, explores a poor action prediction capability and its consequences. In the fourth case a scenario is addressed where two prepared actions exist but one is relatively less positive over the other. For more scenarios, see the Appendix at `http://www.few.vu.nl/~dte220/IWINAC13Appendix.pdf`.

In the example simulations, for the states that are affected by only one state (i.e., in LP1, LP5, LP7, LP8), the function f is taken as the identity function $f(W) = W$, and for the other states f is a combination function based on the logistic threshold function $th(\sigma, \tau, W)$:

$$th(\sigma, \tau, W) = \left(\frac{1}{1 + e^{-\sigma(W-\tau)}} - \frac{1}{1 + e^{\sigma\tau}} \right) (1 + e^{-\sigma\tau}) \qquad (1)$$

Note that the parameter values for the connection strengths used for these scenarios are not unique. Table 4 provides the parameter values used in each case (the changes of weights over cases have been highlighted with a shading color).

Table 3. Specification of Local-Properties

LP1	$SS(W, V_1)$ & $SR(W, V_2) \twoheadrightarrow SR(W, V_2 + \gamma[\, f(\omega_1 V_1) - V_2\,]\,\Delta t)$
LP2	$PA(a, V_1)$ & $PO(a,b,c,s, V_2)$ & $SS(b, V_3)$ & $RAwr(a,b,c,s, V_4)$ & $SR(b, V_5) \twoheadrightarrow SR(b, V_5 + \gamma[\, f(\omega_2 V_1, \omega_3 V_2, \omega_4 V_3, \omega' V_4) - V_5\,]\,\Delta t)$
LP3	$SR(s,V_1)$ & $F(b,V_2)$ & $PAwr(a,b,c,s, V_2)$ & $PA(a, V_3) \twoheadrightarrow PA(a, V_4 + \gamma[\, f(\omega_5 V_1, \omega_6 V_2, \omega_7 V_3) - V_4\,]\,\Delta t)$
LP4	$SR(c,V_1)$ & $SR(s,V_2)$ & $F(b,V_3)$ & $PA(a, V_4)$ & $RO(a,b,c,s, V_5)$ & $PO(a,b,c,s, V_6) \twoheadrightarrow PO(a,b,c,s, V_6 + \gamma[\, f(\omega_8 V_1, \omega_9 V_2, \omega_{10} V_3, \omega_{11} V_4, \omega_{20_6} V_5) - V_6\,]\,\Delta t)$
LP5	$SS(b, V_1)$ & $F(b, V_2) \twoheadrightarrow F(b, V_2 + \gamma[\, f(\omega_{12} V_1) - V_2\,]\,\Delta t)$
LP6	$PO(a,b,c,s, V_1)$ & $PA(a, V_2)$ & $PAwr(a,b,c,s, V_3)$ & $EA(a, V_4) \twoheadrightarrow EA(a, V_4 + \gamma[\, f(\omega_{13} V_1, \omega_{14} V_2, \omega_{15} V_3) - V_4\,]\,\Delta t)$
LP7	$EA(a, V_1)$ & $WS(b, V_2) \twoheadrightarrow WS(b, V_2 + \gamma[\, f(\omega_{16} V_1) - V_2\,]\,\Delta t)$
LP8	$WS(W, V_1)$ & $SS(W, V_2) \twoheadrightarrow SS(W, V_2 + \gamma[\, f(\omega_{17} V_1) - V_2\,]\,\Delta t)$
LP9	$SR(c,V_1)$ & $F(b,V_2)$ & $PO(a,b,c,s, V_3)$ & $EA(a, V_4)$ & $RO(a,b,c,s, V_5) \twoheadrightarrow RO(a,b,c,s, V_5 + \gamma[\, f(\omega_{18} V_1, \omega_{19} V_2, \omega_{20} V_3, \omega_{21} V_4) - V_5\,]\,\Delta t)$
LP10	$RO(a,b,c,s, V_1)$ & $RAwr(a,b,c,s, V_2)$ & $EO(a,b,c,s, V_3) \twoheadrightarrow EO(a,b,c,s, V_3 + \gamma[\, f(\omega_{22} V_1, \omega_{23} V_2) - V_3\,]\,\Delta t)$
LP11	$SR(s, V_1)$ & $F(b, V_2)$ & $PO(a,b,c,s, V_3)$ & $RAwr(a,b,c,s, V_4)$ & $PAwr(a,b,c,s, V_5) \twoheadrightarrow PAwr(a,b,c,s, V_5 + \gamma[\, f(\omega_{24} V_1, \omega_{25} V_2, \omega_{26} V_3, \omega_{27_0} V_4) - V_5\,]\,\Delta t)$
LP12	$PAwr(a,b,c,s, V_1)$ & $F(b, V_2)$ & $RO(a,b,c,s, V_3)$ & $RAwr(a,b,c,s, V_4) \twoheadrightarrow RAwr(a,b,c,s, V_4 + \gamma[\, f(\omega_{27} V_1, \omega_{28} V_2, \omega_{29} V_3) - V_4\,]\,\Delta t)$
LP13	$EA(a, V_1)$ & $WS(s, V_2) \twoheadrightarrow WS(s, V2 + [\, f(\omega^{-} V_1)]\,\Delta t)$

Threshold (τ) and steepness (σ) values used for cases have been listed in Table 5, and particular deviations from that in any case will be stated under that scenario in the respective subsections of Section 4.

4.1 Scenario 1: Normal Execution with Ownership and Awareness

The first scenario considered describes a situation where the context c is the agent itself, and a stimulus s occurs. The action effect b of a, is considered positive for the agent and the awareness of action formation and execution will be scrutinized together with generated prior and retrospective ownership states. The simulation of this scenario is shown in Figure 2. Parameter values used on this can be found in Table 4 under the case 1, and in Table 5. The step size taken is $\Delta t = 0.25$. The slow value 0.5 for γ was applied for external processes modeled by LP6, LP7, and LP8, and the fast value 0.8 for γ for the internal processes modeled by the other LP's.

In Figure 2 it is shown that (after sensing the stimulus) the agent triggers preparation of action a. Based on that the sensory representation of predicted effect b of a is generated (through the as-if body loop) and followed by the feeling of b. Next these states contribute to generate a prior self-ownership. After activating the prior self-ownership, prior self-awareness is developing, mainly upon the formation process of effect prediction b of a. After that, as a result of prior self-awareness and ownership states, the agent initiates the actual execution of action a which propagates its effects through the body loop [13, 14]. In the Figure 2 it clearly shows that the execution of action a (via the body loop) also affects in positive manner via sensoring and the sensory representation of b and the feeling of b. In parallel the sensory representation b of a is suppressed

Table 4. Connection weight values used for cognitive agent model

Weights	Case 1	Case 2	Case 3	Case 4
ω_1	1	1	1	1
ω_2	0.8	0.2	0.4	0.9, 0.8
ω_3	-1	-1	-1	-1
ω_4	0.8	0.8	0.8	1
$\omega_{\cdot}$	-1	-1	-1	-1
ω_5 to ω_{11}	0.8	0.8	0.8	0.8
ω_{20_0}	-1	-1	-1	-1
ω_{12}	0.8	0.2	0.8	0.8, 0.6
ω_{13}	0.9	0.9	0.9	0.9
ω_{14} to ω_{23}	1	1	1	1
ω_{24}	0.8	0.8	0.8	0.8
ω_{25} to ω_{26}	1	1	1	1
ω_{27_0}	-1	-1	-1	-1
ω_{27}	0.8	0.8	0.8	0.8
ω_{28} to ω_{29}	1	1	1	1
$\omega_{\cdot\cdot}$	-0.2	-0.2	-0.2	-0.2

Table 5. Threshold (τ) and Steepness (σ) values used in configurations of simulations

	PA(*a*)	SR(*b*)	PO	PAwr	EA(*a*)	RO	RAwr	EO
τ	0.7	0.2	2.2	1.4	1.5	2.2	1.1	0.9
σ	6	4	7	8	4	12	10	10

due to the prior self-ownership state which causes a dip in the graph (cf. [15, 16]), and that effect will be propagating to the feeling of b too. Due to the action execution the agent develops a retrospective self-ownership state which is followed by a retrospective self-awareness state. At the same time the figure shows that the sensory representation b of a is still getting suppressed due to the retrospective self-awareness (cf. [17, 18]). Finally, the agent communicates self-ownership about the performed action based on retrospective self-awareness and ownership. Note that when the stimulus is taken away, all activation levels will come down to 0 (q.v. LP13), and will come up again when the stimulus reoccurs.

4.2 Scenario 2: Vetoing an Action with No Positive Prediction

The second scenario describes a situation similar to the scenario one, but where the action a triggered by stimulus s has an effect b which is not particularly positive for the agent; here action a hardly has an impact on effect b which would have been positive. The simulation of this scenario is shown in Figure 3. This scenario was modeled by taking the connection strength for the prediction of effect b for action a, and its feeling: ω_2 and ω_{12} as 0.2.

In Figure 3 the predicted effect is very low compared to Scenario 1, though it is having a sufficient sensory representation level. This clearly shows that the action a triggered by stimulus s that has an effect b which is not positive for the agent (in other wards it is more like neutral to the agent in terms of feeling) leads

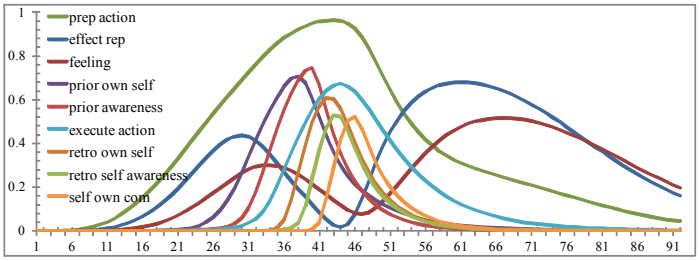


Fig. 2. Scenario 1: Executing an action with ownership and awareness

to not getting any positive feelings out of it. Nevertheless, prediction capabilities are assumed correct in this case, so no high level of b is correctly predicted for a. As a result of this low prediction, the prior self-ownership state also stays at a low level. Due to this, prior awareness is not developed (stays in a very low level), which would be needed to strengthen the action execution. This shows the evidence of facilitating a moderating role of prior awareness (cf. [17, 18]). Therefore execution of the action also stays very low (below 0.1) and due to that, there is no retrospective self-ownership state and nor communication of self-ownership.

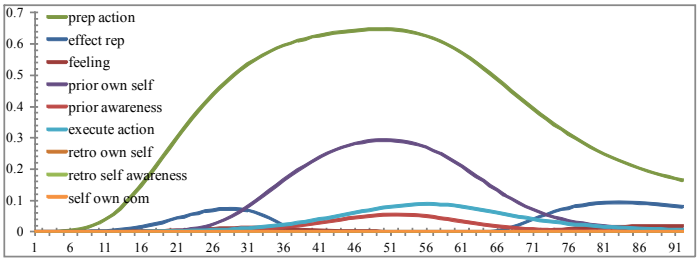


Fig. 3. Scenario 2: Vetoing an action with no positive prediction

4.3 Scenario 3: Effects of Poor Prediction; Schizophrenia Case

The third scenario considered describes a situation where again the context c is the agent itself, and stimulus s occurs. The action effect b for action a, in principle is positive for the agent, like in the first scenario above. However, due to poor prediction capabilities this effect is not (fully) internally predicted. This is what is assumed to happen in patients with Schizophrenia, as discussed, for example, in [7, 8]. The simulation of this scenario is shown in Figure 4. This scenario was modeled by taking the connection strength for the prediction of effect b for action a moderately low: $\omega_2 = 0.2$ (values for the other parameters were again the same as per the scenario 1 and refer Table 4). For this case $\Delta t = 0.5$ was taken instead of 0.25 which was in Scenario 1.

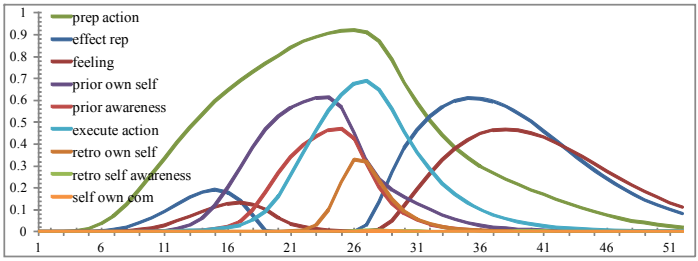


Fig. 4. Scenario 3: Effects of poor prediction (Schizophrenia)

Figure 4 shows that the preparation for action a is followed by the representation of the predicted effect b. The predicted effect is substantially low compared to the first scenario, but higher than in the second scenario. As a result of this predicted effect, prior self-ownership state is developed, and achieves the maximum level of 0.6. Further, in this situation agent is experiencing around a 0.5 level of prior awareness. Therefore, in contrast to the previous scenario, this level turns out high enough for the execution of the action. Nevertheless, only a low level of the retrospective self-ownership state is developed (approximately 0.3), and no retrospective awareness and communication of self-ownership take place, as often happens in patients with Schizophrenia.

4.4 Scenario 4: Executing a Selected Action

The fourth scenario describes a situation where the observed stimulus s leads to two possible action options a_1 and a_2. The action effect b_1 (of a_1) is considered slightly positive for the agent than b_2 (of a_2). Parameter values used can be found in Table 4 under case 4 and in Table 5, except for retrospective self-ownership (τ=2 and σ=12), retrospective self-awareness (τ=0.9 and σ=10) and communication of self-ownership (τ=0.6 and σ=10). Further ω_2 and ω_{12} have two values per each in Table 4 in which the first value is always for a_1 driven execution, and the other for a_2 driven execution. In Figure 5 it is shown that after sensing the observed action preparation for action a_1 and a_2 both starts at the same time with the same gradient in the time interval 6 to 19, and after that its of a_2 getting decreased due to its relatively low sensory representation strength over the a_1. The same phenomenon can be closely examined in the sensory representations of the predicted effects of a_1 and a_2, the feelings of predictions of b_1 and b_2, prior ownership of (a_1, b_1) and (a_2, b_2), and prior awareness of (a_1, b_1) and (a_2, b_2). More importantly the difference of the maximum levels in each group will become high in the above order (0.08, 1.2, and 2.6) while giving the idea of the system is strengthening one option over the other. Therefore; these results contribute to the execution of action a_1 (with the maximum strength level of 0.42) while action a_2 is almost neglected. Also the agent will experience only the retrospective ownership, retrospective awareness and communication attached to the action a_1 as the action a_2 was not performed.

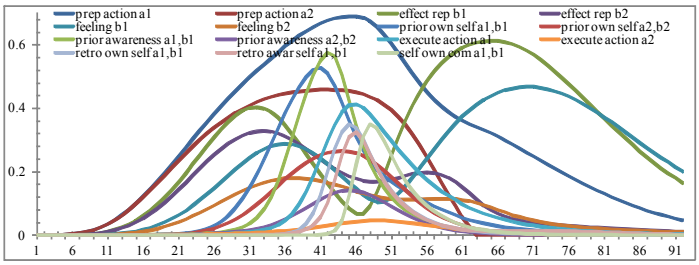


Fig. 5. Scenario 4: Executing a selected action

5 Discussion

The agent model presented here explores how agents prepare for and perform actions and how awareness states can play their role in these processes. Two types of awareness states are distinguished: prior and in retrospect to the execution of the action. The model is a neurologically inspired agent model that is able to make such distinctions. In a number of scenarios it is illustrated how actions are prepared without being conscious initially, but later on awareness states may occur. When prior awareness is developed this may have a decisive effect on actually executing the action, but it might as well be the case that the awareness state has no effect on whether the action is performed. These variations have been illustrated by a wide variety of simulation experiments, some of which were described here; for more scenarios, see: `http://www.few.vu.nl/~dte220/IWINAC13Appendix.pdf`.

The agent model is meant as a basis for subsequent work on developing ambient agent systems able to monitor, analyse and support persons trying to develop a healthy lifestyle. If such systems have such a model of the underlying human processes, they can use this to have a more deep understanding of the human.

In future research it is planned to extend this work by scrutinizing the rationality of the decision making in relation to adaptivity with respect to given environment characteritics. In that respect the role of the feelings generated from internal simulation process and action execution will be considered as well. Furthermore, it may be useful to investigate in more depth the interplay between conscious and non-conscious compounds in human decision making as a regulatory process.

References

1. Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K.: Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): the unconscious initiation of freely volantary act. Brain 106, 623–642 (1983)
2. Haynes, J.D.: Decoding and predicting intentions. Ann. NY Acad. Sci. 1224, 9–21 (2011)

3. Wegner, D.M.: The Illusion of Conscious Will. The MIT Press, Cambridge (2002)
4. Baumeister, R.F., Masicampo, E.J., Vohs, K.D.: Do conscious thoughts cause behavior? Annual Review of Psychology 62, 331–361 (2011)
5. Wolpert, D.M.: Computational approaches to motor control. Trends in Cognitive Sciences 1, 209–216 (1997)
6. Haggard, P.: Human volition: towards a neuroscience of will. Nature Neoroscience Reviews 8, 934–946 (2008)
7. Synofzik, M., Thier, P., Leube, D.T., Schlotterbeck, P., Lindner, A.: Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. Brain 133, 262–271 (2010)
8. Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., Haggard, P.: Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. Brain 133, 3104–3112 (2010)
9. Moore, J., Haggard, P.: Awareness of action: Inference and prediction. Consciousness and Cognition 17, 136–144 (2008)
10. Treur, J.: A computational agent model incorporating prior and retrospective ownership states for actions. Biologically Inspired Cognitive Architectures 2, 54–67 (2012)
11. Damasio, A.R.: Descartes' Error: Emotion, Reason and the Human Brain. Papermac, London (1994)
12. Damasio, A.: The Feeling of What Happens. Body and Emotion in the Making of Con-sciousness. Harcourt Brace, New York (1999)
13. Haggard, P., Eimer, M.: On the relation between brain potentials and the awareness of voluntary movements. Experimental Brain Research 126, 128–133 (1999)
14. Banks, W.P., Isham, E.A.: We infer rather than perceive the moment we decided to act. Psychological Science 20, 17–21 (2009)
15. Blakemore, S.-J., Wolpert, D.M., Frith, C.D.: Why can't you tickle yourself? Neuroreport 11, 11–16 (2000)
16. Fourneret, P., de Vignemont, F., Franck, N., Slachevsky, A., Dubois, B., Jeannerod, M.: Perception of self-generated action in schizophrenia. Cogn. Neuropsychiatry 7, 139–156 (2002)
17. Geldard, F.A., Sherrick, C.E.: The cutaneous "rabbit": a perceptual illusion. Science 178(4057), 178–179 (1972)
18. Dennett, D.C.: Consciousness explained. Little Brown & Co., Boston (1991)
19. Bosse, T., Jonker, C.M., van der Meij, L., Treur, J.: A Language and Environment for Analysis of Dynamics by Simulation. Intern. J. of Artificial Intelligence Tools 16, 435–464 (2007)

Emotional Stress Detection in Contradictory versus Self-consistent Speech by Means of Voice Biometrical Signature

Victoria Rodellar-Biarge, Daniel Palacios-Alonso,
Elena Bartolomé, and Pedro Gómez-Vilda

Grupo de Informática Aplicada al Tratamiento de Señal e Imagen
Neuromorphic Speech Processing Laboratory
Centro de Tecnología Biomédica and Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón,
Madrid Department of Architecture and Technology Systems
Faculty of Computer Science, Polytechnic University, Madrid, Spain
{victoria,pedro}@fi.upm.es, daniel@junipera.datsi.fi.upm.es

Abstract. Most of the parameters proposed for the characterization of the emotion in speech concentrate their attention on phonetic and prosodic features. Our approach goes beyond by trying to relate the biometrical signature of voice with a possible neural activity that might generate voice production. The present study affords emotional differentiation in speech from the behavior of the biomechanical stiffness and cyclicality estimates, indicators of tremor. The emotion under study is the stress produced when a speaker has to defend an idea opposite to his/her thoughts or feelings and compared when his/her speech is self-consistent. The results presented show that females tend to relax vocal folds and decrease tremor and males tend to show the opposite behavior.

Keywords: Speaker's biometry, Glottal Signature, Emotional Tremor, Emotional stress, Phonatory biomechanics Parameters, Voice Production.

1 Introduction

Speech emotion detection could play an important role in Human Computer Interaction in many potential applications as distance learning, testing and usability feedback, sales promotion, older people monitoring, intelligent toys, text/speech converters, language translators, games, costumer centers and forensics, among others. But emotions detection and identification is a very difficult and challenging problem. Since Socrates, philosophers have been concerned about the nature of emotion. Darwin was a pioneer studying how emotions affected the behavior and the language in animals [1]. Rarely emotion has a definition universally used and accepted. The study of emotions is a multidisciplinary

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 74–82, 2013.

field that involves neurological, physiological, psychological, sociological and cultural aspects. Yet today there is not an integrating theory of emotions which has relevance to all areas of concern. Emotions are sometimes mixed and difficult to specify or unravel. Plutchick [2] proposed a model consisting of eight opposite primary emotions: joy/sadness, trust/disgust, fear/anger and surprise/anticipation, whose combinations produce all the emotions normally identified in language. The classification of diverse emotional experiences is described in two orthogonal dimensions: Valence and Arousal. Valence specifies how negative or positive an event is, and Arousal refers to the intensity of an event (neutral or exciting).

The basic steps involved in detect and identify emotions are estimating and selecting the most representative parameters to characterize them and find emotional patterns using classification methods. The key point starts with a basic question, what are the most representative parameters hidden in the sound waves produced by phonatory and articulatory neuromotor actions? Classical literature concentrates its efforts on the analysis of speech sounds from both acoustic and physiological points of view including production and perception (phonetics) and the rhythm, stress and intonation of speech (prosodic). Our approach is based on the idea that emotional states and neurological diseases alter or difficult the precise action of neuromotor activity induced by the brain and produces correlates in voice and speech. The present work is oriented to show how the biometric signature of voice based on the parameterization of the glottal biomechanics and perturbations in the vocal folds stiffness, manifested as a tremor [3], can be used as descriptors of emotions.

2 The Methodology

Speech production starts at the neuromotor cortex which plans the articulatory movements by means of laryngeal nerve activation which controls the vocal folds tension by the enlargement or the shortening of the musculus vocalis. Neuromotor stimuli of the trans- and oblique cricoarytenoid muscles bring both vocal folds together producing a closure of the Larynx. Pressure build-up forces the vocal folds to come apart against viscoelastic muscular forces [4]. The interaction between the glottal flow, the vocal fold adduction and abduction control produces phonation bursts that are filtered by vocal and nasal tracts, and then the speech is produced.

The classical voice production model from Fant [5], shown in Figure 1a), considers that the glottal source is produced by a train of delta pulses $\delta(n)$, neural stimulus, which are modeled by a Glottal Function $F_g(z)$ to reproduce the glottal source u(n). This signal, when is injected in the vocal tract composed by a chain of tubes $F_v(z)$ produces voice $s_v(n)$ which is radiated as s(n). Our approach is inspired in Fant's model and it is carried out by inverting the stages present in speech production to deconstruct it in its components and to relate them to phonation physiology. The method contains three levels of inverse models, as shown in Figure 1b). The first model estimates the glottal source from speech, the second extracts the biomechanical parameters of the vocal folds (dynamic mass, stiffness

and viscosity), and the last one obtains the cyclic parameters from stiffness, which can be an indicator of tremor because they seem to be correlated with the neural control of musculis vocalis. All these parameters are independent and could be unconscious indicators of the emotional state.

The glottal source can be considered as the biometrical signature of voice and can be compared in semantic power with the ElectroCardioGram. As the ECG describes semantic correlates in its singularities, both time and amplitude, the glottogram points to a similar semantics in the biomechanics of the vocal folds. This part begins with the application of an inverse model to compensate lips radiation. Next, an iterative inverse filtering and deconvolution process is used to remove vocal tract effects. The methodology is based on the use of ideal all-pole filters which model the spectral envelopes of the glottal pulse and the vocal tract impulse responses [3].

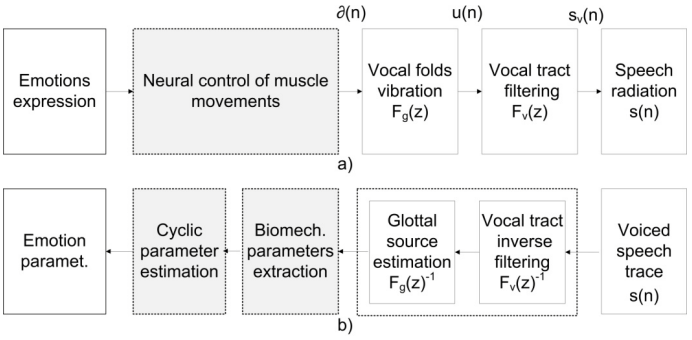


Fig. 1. a) Speech production, b) emotion parameterization from voiced speech trace

The characterization and estimation of the biomechanical parameters of the vocal folds requires the solution of the second inverse problem given the glottal source signal. Figure 2.left) shows a simplified cross section of the vocal folds. They are multilayered structures, consisting of a muscle covered by a mucosal cover. The multilayered structure allows the vocal folds to be stretched or contracted and forced to vibrate at many different lengths. The more internal layer is the thyroarytenoid (vocalis) muscle (body), it runs the entire length of the vocal fold. Surrounding this muscle there is a sheath of mucosal tissue (cover) its stiffness varying from the inner to the outer layer. The stiffness of the mucosa depends on the state of contraction of the laryngeal muscles induced by the air flow coming from the lungs. The interaction between the glottal flow and the vocal folds is a fluid-structure problem, which requires solutions in 3D and time domain. Nevertheless for the purpose of obtaining first-order estimates simpler models may be used reducing the computational complexity of the problem. In this sense, the relations to the body-cover of vocal folds may be modelled as a two mass-spring-damper system [6]. In Figure 2.center), such a model may be seen where M represents the masses and K the elasticities, b: body, c: cover,

l: left vocal fold and r: right vocal fold. The system can be transformed into its electrical equivalent as shown in Fig 2.right), as far as small signal is concerned, it explains the response to external driving forces (f_{cover}, F_{body})), and shows resonance peaks resulting from body-cover mass-spring interactions and in-between valleys induced by inter-elasticity $K_{bcl,r}$. Then, the inverse problem to solve consists in relating the glottal source with the vocal fold vibration estimating the mass μ, viscosity σ and the elastic parameter ξ (or stiffness) of the body and cover biomechanics. This relation is established by minimizing the error between the power spectral density of the residual glottal signal (glottal source signal minus average acoustical signal) and the transfer function of the electromechanical equivalent circuit.

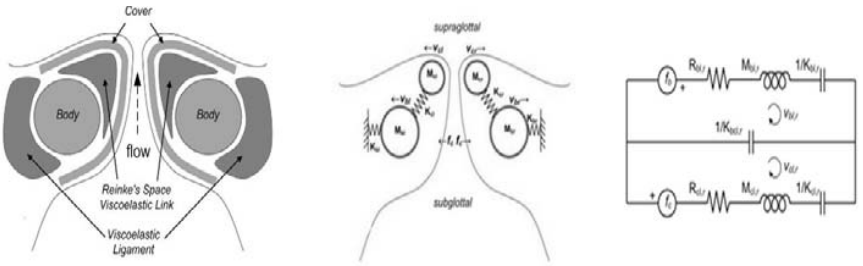


Fig. 2. left) Vocal folds structure, center) two mass model, right) electromechanical equivalent circuit

The results obtained in solving the inverse problem are compliant with fold vibration. In figure 3, stiffness estimates from 20 females and 20 males voice (pathology free) are shown. It can be noticed that both body and cover are larger in females than in males, this observation gives us the hint that any further study must be done separating the population by gender.

The third inversion model is based on the fact that neurological diseases as Parkinson or spasmodic dysphonia produce involuntary voice tremor [7-8] and may cause noticeable voice disorders. Vocal tremor signals are low frequency modulations of voice frequency or amplitude and intermittent voice instability.

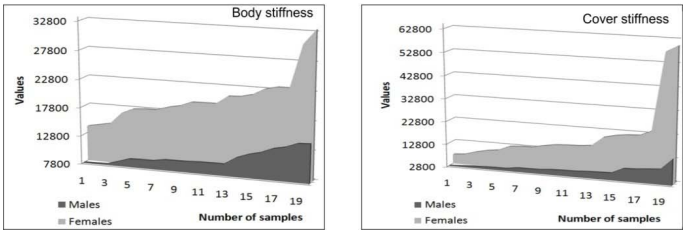


Fig. 3. Body and cover stiffness profiles for male and female

The typical frequency range for noticeable pathological tremor is between 4 and 7 Hz, but frequencies over 10 Hz are not audible perceived. Voice tremor may be associated to respiratory, laryngeal, and articulatory aspects of phonation. Concerning the laryngeal aspect, a deficient neural control of laryngeal muscles may result in vocal tremors. Then, the hypothesis in this work is that emotional states produce also some kind of tremor and it may leave correlates in the biomechanical parameters. Specifically, it is hypothesized that the influence of the emotional alterations have to leave a mark in the stiffness ξ_c on the vocal folds as a cyclic pattern. The estimation of the cyclic behaviour from the obtained stiffness is done by means of an adaptive inverse AR filter [9]. The three lowest-order pivoting coefficients $\{c_{1m}, c_{2m}, c_{3m}\}$ will be used as descriptors of the stiffness cyclic pattern and they are statically independent. These coefficients are pre-normalized to (-1, 1), which allows easier result contrasting. We have seen that c_1 may be an indicator of cyclicality, c_2 and c_3 are used also as co-descriptors, although they do not share the same properties as c_1 [10]. The highest tremor is produced when c_1 achieves values close to -1 and decreases as this parameter increases its value. The ranges of c_1 and c_2 for the same population than in Figure 3 are shown in Figure 4. It must be noticed that c_1 and c_2 are larger in males than in females, which mean, generally speaking, that males present less tremor than females.

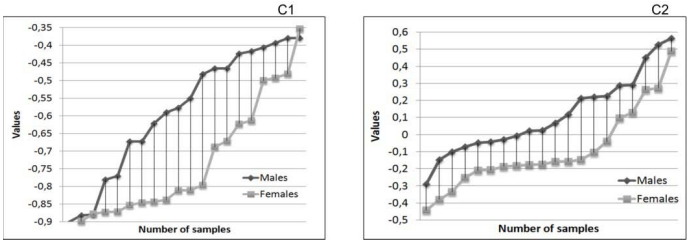


Fig. 4. Ranges of c_1 and c_2 parameters for female and males

3 Self-consistent versus Contradictory Emotions Data Set

Progress in automatic emotions recognition depends heavily on the development of appropriate databases. Emotion expressions and perception are strongly dependent on cultural and sociological aspects and several data bases in different languages have been reported in the literature [11] [12]. But most of them are not available for public use, the nature of its data is very disperse and they do not represent self-consistent emotions. There are speech emotional data collections: acted, elicited and natural. Simulated or acted speech is obtained by asking an actor to speak with a specific emotion. Elicited or non-prompted speech is obtained from provoked emotions in different scenarios [13]. Recording of genuine emotions raises an ethics issue and difficulties with emotional labeling. These recordings are

difficult to handle due to background noise, interferences among speakers, acoustic conditions of the recordings and microphone qualities. Natural speech or real life emotions available data came from call centers, aircraft pilots conversations and media recordings. Acted and natural emotions are labeled after being verified by judges using subjective perception tests. Elicited emotions are labeled with the self-report from the subject under the provoked emotion. The number of subjects in data collections vary from 4 to around 250, the population between males and females is unbalanced, age of the informants is not considered, and data are diverse including vowels, words, complete sentences and records during a fixed time period from each speaker. Most of the databases contain totally or partially the following emotions: anger, boredom, disgust, joy, neutral, sadness and surprise, but there are a very few that include stress. This lack of standardization in databases makes it very difficult to have an scenario for benchmarking parameter quality and classification methods [14]. Then, to have a real progress in emotion detection from speech a big effort must be done in the availability of standardized data that can be shared among researchers [15].

Our speech emotions data set was recorded to evaluate self-consistent versus contradictory or false opinions to elicit emotional stress [16]. Self-consistent speech expresses ideas accordingly to our natural way of thinking and feeling, meanwhile contradictory or false opinions require to artificially fabricate them and to prepare the speech. It might produce different marks as disfluency, repetitions, delays, fillers and longer vowels on the speech, among others. Spontaneity is lost and it produces stress, which implicates physiological alterations on the speaker. Our recordings were made under the following conditions: the informants should not have any alteration in speech production or perception and they must be native Spanish speakers. The data collection has 80 samples gender balanced. Age, social class, cultural level and dialect had not been considered. The procedure to obtain the data was the proposed by Arciuli [17] consisting in two interviews. In the first interview the valence and the arousal were evaluated. Each informant was given a very controversial social topics questionnaire with seven questions, as: "the crisis in Spain is due to politicians?", "man should work, woman should stay at home?","is it fair that you continue paying your mortgage after the bank has taken your house?", etcetera. Then each informant had to give their opinion about each topic grading it from 1 to 7 (strongly disagree = 1, strongly agree = 7) and they have also to grade their feelings about each topic with the same scale (indifference = 1, very strong feelings = 7). Then two topics from each informant were selected, those which presented both a strong opinion/valence (agreement or disagreement) and strong feelings/arousal. In the second interview, informants are asked to defend their self-consistent opinions in one of the topics and to defend their contradictory opinion in the other, each answer must be expressed in 20 seconds. Later on, samples of maintained vowels /a, e, o/ in modal phonation from each informant are also recorded. The speech recording is done by a different person than the one who made the first interview and he does not know when the informant is lying o telling the truth, then informants can express their opinions more self-consistently to convince the

interviewer that they are telling the truth. The difficulty to artificially construct and express a contradictory opinion in only 20 seconds reflects an increment of stops and fillers pronounced inadvertently on the speech. The most common filler used by Spanish speakers is vowel /e/ from words as /de/ and /que/. Those fragments were isolated and used in this study.

4 Results and Discussion

Biomechanical parameters (vocal fold body and cover mass stiffness) and tremor cyclicality (c_1 and c_2) were obtained from 400 ms long segment from /eh/ fillers from the 20 females and 20 males of our data collection. Comparisons between pre- (self-consistent) and post- (contradictory) estimates from the same subject, will be presented next. The comparisons are based on the observed variations in the parameters for contradictory versus self-consistent speech. This variations have been coded as follows: H means an increment in the value of the parameter in contradictory speech, L means a decrement, and S a non-significant variation. The values below 20% of the mean of the contradictory minus self-consistent values, were labeled as S.

In Table 1 the results for females are shown. It may observed, that the estimated values of body stiffness have suffered a decrement on fourteen women (70%), an increment on three women (15%), and it remains about the same value for the other three (15%). The cover stiffness has the same tendency than body stiffness, and it shows an decrement on eleven cases (55%), an increment on seven cases (35%) and no significant changes on two cases (10%). c_1 and c_2 present the opposite behavior than stiffness. C_1 is incremented on fifteen cases (75%) and it decreases on five (25%). While c_2 is incremented on twelve cases (60%) and decreases in eight (40%). From these results we can infer that stress in most of the females implicate a decrement on the body and cover stiffness and an increment on the cyclic parameters. Then it seems that, their tendency is to relax the vocal folds and to decrease tremor.

Table 1. Parameter variations for Females

Woman	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Body	H	**L**	**L**	**L**	**L**	**L**	**L**	S	H	**L**	**L**	S	S	**L**	**L**	H	**L**	**L**	**L**	L
Cover	H	**L**	H	H	**L**	**L**	S	H	H	**L**	**L**	**L**	**L**	H	**L**	H	S	**L**	**L**	**L**
C1	**H**	**H**	**H**	**H**	**H**	**H**	L	L	L	**H**	**H**	**H**	**H**	**H**	**H**	**H**	L	**H**	**H**	L
C2	**H**	**H**	L	**H**	**H**	**H**	L	L	L	**H**	**H**	**H**	L	**H**	**H**	L	L	**H**	**H**	L

The results for men are shown in Table 2, and as we will see indicate an opposite behavior than the ones obtained for women in Table 1. In men, the body stiffness values increase in 60% of cases, they decrease in 30% of them, and have similar values in the rest 10%. The cover stiffness increases also in 50% of cases, while decreases in 45% and remains in 5%. The first cyclic parameter

Table 2. Parameter variations for Males

Man	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Body	**H**	**H**	**H**	**H**	L	L	**H**	**H**	S	L	**H**	L	S	L	**H**	**H**	**H**	**H**	**H**	L
Cover	**H**	**H**	S	**H**	L	**H**	**H**	**H**	**H**	L	L	L	**H**	L	L	**H**	**H**	L	L	L
C1	H	**L**	**L**	**L**	**L**	**L**	**L**	H	**L**	**L**	H	H	H	**L**	**L**	H	**L**	**L**	**L**	H
C2	H	**L**	**L**	**L**	H	H	**L**	**L**	**L**	H	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	H

decreases in a percentage of 65% and it increases in the rest 35% of the cases. And finally the second cyclic parameter has the same behavior than the first one, where 75% of the cases experiment an increase and 25% a decrease. So under the emotional stress described, males tendency is to increase the values of stiffness, as a consequence to tense vocal folds, and to decrease c_1 and c_2 increasing the tremor.

In a first approach, the proposed biological parameters seem to be suitable indicators to characterize emotional stress in speech. In female, the results reflect a variability 85% in body stiffness and 90% in cover stiffness. In male, the body and cover stiffness present 5% more of variability than female. In both genders the variability is a 100% in the cyclic parameters. Nevertheless the tendency in the parameter variability (increment or decrement) seems to be more clearly defined in females than in males. The difficulty to obtain reliable data is the major problem in this research, data samples cannot be taken a second time from the same speaker, because spontaneity is lost. And on the other hand, the questions asked are very committed, therefore some people may answer according what is politically correct instead of their self-consistent opinions and feelings.

5 Conclusions

The main problem of emotion detection and identification is the difficulty to obtain self-consistent reliable data but this problem is extremely difficult to solve. In spite that there are some acted, elicited and natural database available, they lack enough standardization to become benchmarks to make some progress in the field. Most of the parameters proposed for the characterization of the emotion hidden in speech focus their attention on the phonetic and prosodic features, which are easily impostable. Our approach goes an step further by trying to relate the biometrical signature of voice with the neural activity that might generate voice production. We have studied stress caused by forcing to a person to express non-self-consistently and self-consistently with their opinions. Our results show that males and females have an opposite behavior under this kind of stress. Females tend to relax the vocal folds while males tend to tense them, this fact leaves a mark in the parameter related with the tremor as well. Future research will be oriented to study EGC responses in relation with the research presented in this paper.

Acknowledgments. This work has been funded by grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Science and Technology, Spain.

References

1. Darwin, C.: The Expression of the Emotions in Man and Animals (1872)
2. Plutchick, R.: The Psychology and Biology of Emotion. Harper Collins Publishers (1994)
3. Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L.M., Martínez, R., Godino, J.I.: Glottal Source Biometrical Signature for Voice Pathology Detection. Speech Comm. 51, 759–781 (2009)
4. Luschei, E.S., Ramig, L.O., Baker, K.L., Smith, M.E.: Discharge characteristics of laryngeal single motor units during phonation in young and older adults and in persons with Parkinson disease. J. Neurophysiol. 81, 2131–2139 (1999)
5. Fant, G.: Acoustic Theory of Speech Production, 2nd edn. Mouton (1970)
6. Story, B., Titze, I.: Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. 97(2), 1249–1260 (1995)
7. Barkmeier, J., Hoffman, H.: Synchrony of laryngeal muscle activity in persons with vocal tremor. Arch. Otolaryngol. Head Neck Surg. 129(3), 313–323 (2003)
8. Pérez, K., Ramig, L., Smith, M., Dromey, C.: The Parkinson larynx: tremor and videostroboscopic findings. J. Voice 10(4), 354–361 (1996)
9. Deller, J.R., Proakis, J.G., Hansen, J.H.L.: Discrete-Time Processing of Speech Signals. MacMillan Pub. Co., Englewood Cliffs (1993)
10. Rodellar,V., Palacios, D., Bartolomé, E., Gómez, P.: Vocal Fold Stiffness Estimates for Emotion Description in Speech. Proceedings of BIOSIGNALS 2013 – 6th International Conference on Bio-inspired Systems and Signal Processing, pp. 112-119 (2013)
11. Ververidis, D., Kotropoulos, C.: A Review of Emotional Speech Databases. In: Proc. 9th Panhellenic Conference on Informatics (PCI), Thessaloniki, Greece, pp. 560–574 (2003)
12. Ramakrishnan, S.: Recognition of Emotion from Speech: A Review. In: Speech Enhancement, Modeling and Recognition-Algorithms and Applications. InTech (2012)
13. Ververidis, D., Kotsia, I., Kotropoulos, C., Pitas, I.: Multi-modal emotion-related data collection within a virtual earthquake emulator. In: Proc. Inter. Conf. Language Resources and Evaluation (LREC), Morocco (2008)
14. Hasrul, M., Hariharan, M., Sazali, Y.: Human Affective (Emotion) Behavior Analysis using Speech Signals: A Review. In: Proceedings of the 2012 International Conference on Biomedical Engineering, ICoBE (February 2012)
15. Bhutekar, S., Chandak, M.: Designing and Recording Emotional Speech Databases. International Journal of Computer Applications (IJCA): Proceedings on National Conference on Innovative Paradigms in Engineering and Technology (NCIPET 2012) 4, 6–10 (2012)
16. Frank, M.G., Ekman, P.: Appearing truthful generalizes across different deception situations. Journal of Personality and Social Psychology 86, 486–495 (2004)
17. Arciuli, J., Villar, G., Mallard, D.: Lies, Lies and More Lies. In: 31st Annual Meeting of the Cognitive Science Society (CogSci 2009), pp. 2329–2334 (2009)

Vision Substitution Experiments with See ColOr

Guido Bologna, Juan Diego Gomez, and Thierry Pun

Computer Vision and Multimedia Lab, University of Geneva, Switzerland
{Guido.Bologna,Juan.Gomez,Thierry.Pun}@unige.ch
cvml.unige.ch

Abstract. See ColOr is a mobility aid for visually impaired people that uses the auditory channel to represent portions of captured images in real time. A distinctive feature of the See ColOr interface is the simultaneous coding of colour and depth. Four main modules were developed, in order to replicate a number of mechanisms present in the human visual systems. In this work, we first present the main experiments carried out in the first years of the project; among them : the avoidance of obstacles, the recognition and localization of objects, the detection of edges and the identification of coloured targets. Finally, we introduce new undergoing experiments in Colombia with blind persons, whose purpose is (1) to determine and to touch a target; (2) to navigate and to find a person; and (3) to find particular objects. Preliminary results illustrate encouraging results.

Keywords: 3D-vision, vision substitution, colour-depth sonification, human-computer interaction.

1 Introduction

The purpose of the See ColOr project is to create a mobility assistance device for visually impaired people, in order to make a step further toward their independent mobility. See ColOr is a non-invasive mobility aid for blind users that uses the auditory pathway to represent in real-time frontal image scenes. The basic idea is to encode a number of coloured pixels by spatialised musical instrument sounds, in order to represent and emphasize the colour and location of visual entities in their environment. Specifically, the points captured by a camera are represented as directional sound sources, with each emitted sound depending on the colour of the pixel. Finally, pixel depth has been represented by sound rhythm.

Several authors proposed special devices for visual substitution by the auditory pathway in the context of real time navigation. The K Sonar-Cane combines a cane and a torch with ultrasounds [9]. Note that with this special cane, it is possible to perceive the environment by listening to a sound coding the distance. "TheVoice" is another experimental vision substitution system that uses auditory feedback. An image is represented by 64 columns of 64 pixels [11]. Every image is processed from left to right and each column is listened to for about 15 ms. Specifically, every pixel gray level in a column is represented by a sinusoidal

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 83–93, 2013.

wave with a distinct frequency. More recently, "EyeMusic" sonifies a 24x40 pixel coloured image over 2 seconds [10]. Each colour (red, green, blue, yellow, white, black) is encoded through timbre (e.g. White = piano, Blue = marimba, etc.). The colours vertical location is denoted through the pitch, using notes across 8 octaves of a pentatonic scale. The luminance of each colour is coded into the loudness of the note and similarly to "TheVoice", each column is sequentially presented over time.

In in this work we summarize the main experiments performed with the See ColOr interface connected to a camera. Moreover, we present new undergoing experiments and their preliminary results obtained with blind participants. In the following sections we present the See Color mobility aid, then we illustrate experiments performed in the past with particular modules of our interface, the new undergoing experiments, followed by a discussion and a conclusion.

2 The See ColOr Mobility Aid

2.1 See ColOr Layout

The See ColOr mobility aid architecture is composed of four main modules aiming at replicating several process involved in the human visual system. These modules are :

- the local perception module;
- the global perception module;
- the alerting system;
- the recognition module.

Figure 1 illustrates a user with a See ColOr prototype. Images are captured by the Kinect camera (Microsoft). The *local perception module* presents the user with a row of 25 points in the centre of an image frame.

These points are coded into left-right spatialised musical instrument sounds, in order to represent and emphasize the colour and location of visual entities in the environment (see Sect. 2.2 for the sound code). Sonifying the visual data in a small central part of the image introduces the tunneling vision phenomenon; an important drawback that greatly diminishes the user perception.

To rectify this deficiency, we introduced the *global module*. It includes a multitouch interface (embedded in a tablet) that allows the user to compare different points of a scene. Hence, this interface makes it possible to explore the current scene globally and more proactively; i.e. the image is made entirely accessible. In general, when the user wants to explore the nearby space, he/she will rapidly scan the touchpad with one or more fingers; the principle here is that finger movements replace eye movements and more information from the image as a whole is made accessible all at once. In this way, the user will not have to move his/her head. The finger contact point activates a sound that transmits the corresponding data representation (colour and depth). The spatial position

Fig. 1. The See ColOr prototype

of each finger on the touchpad is spatialised on the azimuth plane but not on elevation. We assume that the user is aware of the vertical position of the touched points.

The purpose of the *alerting system* is to warn the user whenever an obstacle is potentially in his/her trajectory. Roughly, when a cluster of points in the video presenting a distance below one meter continues to approach over a given number of frames, the user should stop to walk. Note also that the alerting system runs simultaneously with respect to local or global modules, so that the user will focus on the exploration.

The *recognition module* is based on the approach proposed by Kalal et al. [8]. We implemented a detecting-and-tracking hybrid method for learning the appearance of natural objects in unconstrained video streams. The object of interest is manually defined in a single frame. Then, we simultaneously follow the object and learn a detector to distinguish its appearance (frame by frame) despite rotations, partial occlusions, and/or perspective changes across a training sequence. Opposite to the authors in [8], we split up the technique in two independent tasks: offline, yet real time training and online detecting-based navigation [7].

2.2 Sound Code

A pixel is represented in the HSL colour space by three variables : hue, saturation and luminosity. The following rules are applied :

- Hue is specified using a musical instrument timbre for each colour: Red (Oboe), Orange (Viola), Yellow (Pizzicato), Green (Flute), Cyan (Trumpet), Blue (Piano) and Purple (Saxophone). Between category hues mix the volume of the neighbouring hue-timbres, allowing smooth transitions.
- Saturation is quantised through the pitch of the hues instrument: $0-0.24$ (C note), $0.25-0.49$ (G note), $0.5-0.75$ (B flat note) and $0.75-1$ saturation (E note).
- Luminance is coded through the addition of either a double bass (< 0.5 luminance) or a singing voice (> 0.5 luminance) to the sound mix, which are played in one of the following notes (in order of increasing brightness); C (darkest variant), G, B-flat or E (brightest variant).

In addition, when luminance values are near 0, whatever the hue/saturation values, only the double bass is played. Similarly, when luminance is almost maximal only the singing voice is selected. The spatial distribution of the 25 sonified pixels is coded into inter-aural time delays and intensity differences between the ears that are used in natural sound localization. Finally, depth is conveyed through the length of sounds, between 90 and 300ms. Note that the sound is briefer for close depth.

3 Experiments Performed in the First Years of the See ColOr Project

Only the first three modules were tested in the first years of the project. Video experiments are available on http://www.youtube.com/guidobologna. Note that a part of the recognition module has been implemented very recently [7] and tested in new experiments with blind people in Colombia.

3.1 Experiments with the Local Module

The first experiment with the local module focused on the matching of coloured sock pairs [1]. Note that in this experiment the purpose was not to recognize colours, since this would have required more training sessions. Basically, experiment participants were trained on musical instrument sounds and also on how to correctly point the camera on socks. Overall, participants matched ten sock pairs with an accuracy of 94%. This experiment demonstrated that blinfolded individuals were able to manipulate objects by pointing a camera on them and also that colours can be matched with high accuracy, even after a short training phase.

The purpose of the first mobility experiment was to follow a red serpentine path painted on the ground for more than 80 m [2]. Ten blindfolded individuals were trained approximately for ten minutes on the recognition of the related sound pattern. On average, all the experiment participants followed the red path at an average speed of 0.74 m/s. Finally, a blind person also succeeded in that experiment.

Later, we performed other mobility experiments, but only with a very well trained person [3]. In these experiments, for the first time colour and depth were sonified, simultaneously. The main results of the experiments were that with the use of a stereoscopic camera it was possible to :

- Detect an open door in order to go out from the office.
- Walk in a hallway and look for a blue cabinet.
- Walk in a hallway and look for a red tee shirt.
- Avoid two red obstacles and pass trough the space separating them.
- Move outside and to avoid a parked car.

The experiment participant was able to avoid big obstacles (walls, doors, cars, etc.) and to find big coloured objects. Generally, the detection of small obstacles or small objects is a much more difficult task. The main reason is that the local module detects only a small part of a video frame. Thus, a user would have to span the current scene many times, before being able to determine small obstacles.

3.2 Experiments with the Global Module

The global perception module aims at providing the user with a tool to compare the characteristics (colour, distance, position) of the main components of a picture. Typically, a scene is represented on a tablet. When the user touches its surface with one or more fingers, she/he receives one or more sounds (simultaneously) representing colours and depths of the contact points. Moreover, sounds are spatialised in the azimuth plane. For instance, if a finger is in the left part of the tablet the sound is listened to the left. Currently, the number of fingers is limited to two.

Object Recognition and Localization with the Use of Depth Maps. The basic idea of this experiment was to provide the user with a depth map that makes it possible to identify and localize objects. As an example, Figure 2 illustrates an example of an experiment participant exploring a scene on a tablet that represents three geometrical objects lying on a table.

A computer-vision-based strategy was implemented to recognize simple geometrical objects present in a real time scene [5]. Afterwards, these objects were associated to sounds of musical instruments and sonified through. Users were able to hear, explore and understand the scenery composition, while finger-triggering the sounds within the tablet surface on which a top-view image of the scene was mapped.

The results revealed that the recruited ten participants were capable of grasping general spatial structure of the sonified environments and accurately estimate scene layouts. The exploration time instead, varied according to the number of elements on the table. In average for a scene composed of three elements, 3.4 minutes were enough to build its layout in mind, whereas for scenes with four elements this time reached 5.4 minutes. This difference was given due to the increase in the number of sound-colours associations to be learned; the results showed no misclassifications of objects, though.

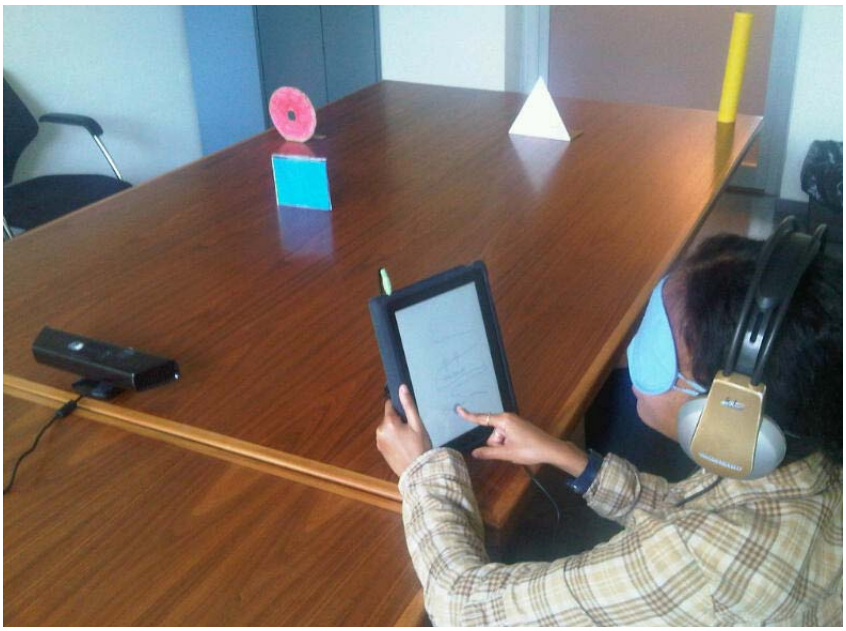


Fig. 2. A blindfolded individual trying to identify and to locate three different objects with an iPad

One or Two Fingers? We performed another experiment for which the purpose was to determine the number of fingers needed for efficient exploration of coloured squares [4]. In the experiments a group of 20 blindfolded users was asked to find colour matches into an image grid represented on a tablet by listening to their associated colour-sound representation. We defined images of coloured squares of increasing difficulty, such as 3x3 and 4x4 grid squares. Figure 3 illustrates a participant performing this experiment with a tactile tablet. The results showed that for the easiest level of difficulty (2x2 squares) the use of two fingers is more efficient than the use of one finger. However, contrary to our intuition, this cannot be statistically confirmed in similar tasks of increasing difficulty.

Edge Detection, Target Identification, and Wall Detection. We recruited 12 individuals for the following three case studies [6] :

- Edge detection.
- Target identification.
- Wall detection.

As an example, Figure 4 illustrates two participants in two experiments.

The experiment on edge detections concerned the capacity of the users to perceive, through the audio feedback, points in an image at which its aspect

Fig. 3. A blindfolded individual looking for squares of the same colour on a tactile tablet

changed sharply or, more formally, had discontinuities. The global module, as well as the local module allows the perception of two classes of edges; those defined by depth variations (e.g. an open door frame), as well as the ones caused by colour interruption (e.g. the boundary of a coloured painting hanging on a white wall). After training, participants were tasked to identify five borders in each case : variation of sound timbre (colour-caused edges) and rhythm of the sound (depth-caused edges). They were blindfolded and asked to answer whether or not at least one edge was present on the tablet. In the affirmative cases, they had to prove their answer by sliding their finger along the edge. The results showed that edge detection was perfectly achieved for colour, with less than 10% errors for depth.

The experiment on the identification of a coloured target aimed at evaluating capabilities of individuals to seek and find a specific target disposed into the environment being explored. We substituted the gaze scanning of a sighted individual by a global exploration with the fingers on a tablet. During the experiment, a person acting as a target and wearing a red t-shirt was in front of the blindfolded user at an unknown location. The blindfolded person was then tasked to explore with his fingers the whole panorama being captured by the camera and rendered onto the tablet. We fixed a controlled environment so that no red elements others than the t-shirt were present. This simplified the task into detecting which portion of the iPad screen emitted the sound of red when touched. This trial was run five times as the target person moved arbitrarily. The results confirmed that target detection was perfectly achieved by all individuals.

Fig. 4. Experiments with the global module. (a) Finding a target, (b) detecting and then approaching safely a wall.

The experiment on adjacent wall detection concerned the ability to be aware of oneself in space. Specifically, in this study we investigated awareness of spatial relationships as the skill to perceive an object in relation to oneself. For example, a person with spatial awareness understood that as (s)he walked towards a door, the door was becoming closer to his/her own body.

Blindfolded participants after a short training phase were able to accurately identify the rhythm of the sound that a wall produces in our system according to its nearness. Afterwards, they were five times asked to walk down the corridor (from different distances) toward a wall and stop right before a hit, but close enough to reach it by hand. The results showed that approximately 10% of the participants required our intervention not to hit the wall.

3.3 Experiments with the Alerting System

The task of detecting objects lying on the user way and on which he or she is likely to stumble, relies on range images. We defined a layer of riskiness within the range of the depth camera. This layer is fixed at 0.9 m and we scan it constantly to determine whether or not, it remains clean of objects. Once we detect a number of connected pixels larger than a certain tolerance, we say that something is on the way. This entity already detected into a plane 0.9 m in front of the user, might very well not be an obstacle, depending on the actual trajectory of the user. For example, the frame of an opened door might not be an obstacle (even if detected 0.9 m in front) provided that the user is just squeezing through the opened door. In that case, we must be able to conclude that this frame will pass on the user side and a stumble is likely not to happen. To achieve this goal, we developed two strategies.

In our first approach once the object is detected in the riskiness layer, we recover the last 15 frames back in time before the detection. Therefore, we can estimate the trajectory of the detected object within those 15 frames (1 second before detected). If we assume that this trajectory is not likely to change abruptly

within the next second, we can predict the final position of the object after traveling from the riskiness layer up to the user layer (1 second after detected).

In our second approach we constrained the detection into the riskiness layer, only to the area that defines the path the user is actually walking on and not to the sideway areas of the scene. Thus, only objects truly lying in front of the user (not in a layer in front) are detected. This allows us to restrict the video into the target area and kept the sideway parts of the scene out of shoot. The second was faster and more likely to run in real time. A demonstration video can be watched on http://youtu.be/X426HAZaiYQ

4 New Experiments

Recently, we started new experiments with blind individuals based on a collaboration with a Colombian government initiative towards work inclusion for individuals with disabilities ("Pacto para la productividad" in Pereira Risaralda). Two of these experiments are depicted in Figure 5. We are planning to have more participants. Here, we describe the experiments carried out by two of them. Headphones have been replaced by bonephones; this is particularly well appreciated by blind persons, as they are able to listen to natural sounds. Note that a short training phase is performed before measuring the performance in the tests.

Fig. 5. Two new experiments. (a) Locating and reaching red target from a spinning chair, (b) Finding a person in a room.

In the first experiment, participants sitting on a spinning chair have to locate a coloured red target with the use of the local module. Then, they have to stand up and to touch it. The alerting system was activated in order to avoid to hit the walls or the target. Our two participants were asked to repeat the experiment four times. At the beginning of a trial the role of the spinning chair is to confuse the user, so as to forget the previous target position (if any). The results showed that the participants were both successful and their average time to reach the target was 3.2 and 2.6 minutes, respectively.

In the second experiment the purpose is to find a particular person in an empty room of approximately 20 squared meters. We use the recognition module that was trained to recognize a particular face. For each participant the experiment was repeated twice. Each time the person to be found changed his place, randomly. Again, the alerting system was used to avoid to be too close to the walls. Our two participants successfully achieved this task with an average time of 5.6 and 6.2 minutes, respectively.

Finally, the last experiment is related to finding particular objects on a table. As in the previous experiment, our participants relied on the recognition module. The objects to recognize are a cap and a remote control. With each object the experiment was repeated three times. The average time to find the cap was for our two participants 4.2 and 5.0 minutes, respectively. Finally, for the remote control the average time was 5.2 and 5.9 minutes, respectively.

5 Discussion and Conclusion

In this work we presented the main experiments performed in the last 6-7 years of the See ColOr project. Almost all the participants were not blind. Actually, we are in the process of achieving new experiments with ten visually impaired persons; this represents a very important step to validate our prototype. Finally, the preliminary results with two participants are very encouraging with respect to the success rate, though we would like to observe faster execution time. In the future we will complete the statistics of the recent experiments.

Aknowledgements. We would like to thank the Swiss National Foundation for funding this research under grant 200020_140392/1. We also would like to thank the Hasler Foundation for the grant that allowed us to start this research between 2005 and 2008. Finally, we thank Dr. Constanza Lopez, head of the project "Pacto para la productividad" in Pereira Risaralda; a Colombian government initiative towards work inclusion for individuals with disabilities.

References

1. Bologna, G., Deville, B., Vinckenbosch, M., Pun, T.: A perceptual interface for vision Ssubstitution in a color matching experiment. In: Proc. Int. Joint Conf. Neural Networks, Part of IEEE World Congress on Computational Intelligence, Hong Kong, June 1-6, pp. 1621–1628 (2008)
2. Bologna, G., Deville, B., Pun, T.: Blind navigation along a sinuous path by means of the See ColOr interface. In: Mira, J., Ferrández, J.M., Álvarez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2009, Part II. LNCS, vol. 5602, pp. 235–243. Springer, Heidelberg (2009)
3. Bologna, G., Deville, B., Gomez, J.D., Pun, T.: Toward local and global perception modules for vision substitution. Neurocomputing 74(8), 1182–1190 (2011)
4. Gomez, J.D., Bologna, G., Deville, B., Pun, T.: Multisource sonification for visual substitution in an auditory memory game: one or two fingers? In: Proc. of the 17th International Conference on Auditory Display, Budapest, Hungary, June 20-23, pp. 1529–1534 (2011)

5. Gomez, J.D., Mohammed, S., Bologna, G., Pun, T.: 3D Scene accesibility for the blind via auditory-multitouch interfaces. In: Proc. of the AEGIS Workshop and International Conference (Accessibility Reaching Everywhere), Brussells, Belgium, November 28-30 (2011)
6. Gomez, J.D., Bologna, G., Pun, T.: Spatial awareness and intelligibility for the blind: audio-touch interfaces. In: Proc. of CHI 2012 (International Conference on Human-Computer Interaction), Austin, Texas, US, May 5-10, pp. 1529–1534 (2012)
7. Gomez, J.D., Bologna, G., Pun, T.: Non-visual-cueing-based sensing and understanding. In: Proc. of the 14th International ACM SIGACCESS Conference on Computers Accessibility (ASSETS 2012), Boulder Colorado, U.S, October 22-24, pp. 213–214 (2012)
8. Kalal, Z., Matas, J., Mikolajczyk, K.: P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In: Proc. of CVPR 2010, International Conference on Computer Vision and Pattern Recognition, San Francisco, CA, US, June 13-18, pp. 49–56 (2010)
9. Kay, L.: A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. The Radio and Electronic Engineer 44, 605–627 (1974)
10. Levy-Tzekdek, S., Hanassy, S., Abboud, S., Maidenbaum, S., Amedi, A.: Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. Restroative Neurology and Neuroscience 31, 1–11 (2012)
11. Meijer, P.B.L.: An experimental system for auditory image representations. IEEE Trans. Bio. Eng. 39(2), 112–121 (1992)

Robot Cognitive Stimulation for the Elderly

Adriana Tapus[1] and Alexandra-Maria Vieru[2]

[1] Robotics and Computer Vision Lab, ENSTA-ParisTech, 828 Blvd. des Marechaux, 91120, Palaiseau, France
adriana.tapus@ensta-paristech.fr
[2] Politehnica University of Bucharest, Bucharest, Romania
alexandra.vieru@cti.pub.ro

Abstract. In this paper, we present a new social intelligent robotic system used for cognitive stimulation therapy for individuals suffering from Mild Cognitive Impairment (MCI) and/or Alzheimer's disease. Our work aims increasing the cognitive attention of users by playing different games and therefore help slow down cognitive decline. The system records the users' task performance during the games and adapts the different levels of difficulty as a function of the users' game history. The games were tailored to the needs of each individual so as to address their different cognitive disabilities. Two studies are presented.

1 Introduction

The life span of ordinary people is increasing steadily and many developed countries, within and outside Europe, are facing the big challenge of dealing with an ageing population at greater risk for cognitive disorders. Mild Cognitive Impairments (MCI) [1] is considered to be the transitional stage between normal ageing and dementia. The age-related decline of cognitive functions generally refers to a mild deterioration in memory performance, executive functions, and speed of cognitive processing. Various brain exercises are used by therapists so as to maintain cognitive functions. These cognitive exercises require a trained therapist to guide the individual through their execution, to design a new configuration, to provide an useful feedback during the task, and to keep track of the user's performance history in order to draw a conclusion on his/her evolution over time. However, space and staff shortages are already becoming an issue, as the elder population continues to grow.

Robotic systems are now capable of social interaction with human users, presenting a new opportunity for providing individualized care and ameliorate their quality of life. A great deal of attention and research is dedicated to assistive systems aimed at promoting ageing-in-place, facilitating living independently in one's own home as long as possible and cognitively stimulating the users.

To the best of our knowledge, very little long-term research has been done in the area of therapeutic robots for individuals suffering from dementia and mild cognitive impairment (MCI). In [5], the authors showed that the presence of a human-like robot whose gestures and speech were human-like, could increase the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 94–102, 2013.

interest of the user in the execution of his/her tasks. The feedback of the persons interacting with a more relational robot (i.e., praises, reassuring feedback in case of failure, continuity in speech) was found more positive than the one received from the persons who interacted with a less communicative robot. The intrinsic motivation was higher at the former due to the social behavior of the robot. Moreover, the intrinsic motivation of individuals was also increased if the robot encourages them to perform better giving as argument their higher scores until present. Also, the users were more motivated when they had the ability to choose the type of exercises to execute. Another experiment is depicted in [8]. This study showed that the use of the person's name, the eye-contact, and a friendly attitude towards the person were more appreciated by the users. In their research work [10], Libin and Cohen-Mansfield describe a preliminary study, which compares the benefits of a robotic cat and a plush toy cat as interventions for elderly persons with dementia. Furthermore, Kidd, Taggart, and Turkle [9] use Paro, a seal robot, to explore the role of the robot in the improvement of conversation and interaction in a group. Marti, Giusti, and Bacigalupo [11] justify a non-pharmacological therapeutic approach to the treatment of dementia that focuses on social context, motivation, and engagement by encouraging and facilitating non-verbal communication during the therapeutic intervention. Moreover, in [15] a new adaptive robotic system based on the socially assistive robotics (SAR) [6] technology that tries to provide a customized help protocol through motivation, encouragements, and companionship to users suffering from cognitive changes related to aging and/or Alzheimer's disease is presented.

Nonpharmacological treatments focus on physical, emotional, and mental activity. Engagement in activities is one of the key elements of cognitive impairments care [13]. Activities (e.g., music therapy [3], [15], [7], arts and crafts) help individuals with dementia and cognitive impairment maintain their functional abilities and can enhance quality of life. Other cognitive rehabilitation therapies and protocols focus on recovering and/or maintaining cognitive abilities such as memory, orientation, and communication skills.

We propose an integrative research work that brings together interdisciplinary expertise from engineering, gerontology, psychiatry, and cognitive and social sciences with a specific target user population. The cognitive stimulation exercises/games will focus on different cognitive functions, especially on:

- **Attention:** It is the ability to process stimulus: for a period of time (sustained attention), to process relevant stimulus but not irrelevant ones (selective attention) and to process more than one stimuli (divided attention). Age-related decline is found in these three kinds of attention.
- **Memory:** It is the capacity of store and use processed information. The ability to process the information already stored and/or to be stored (working memory), which is very relevant for daily life, decreases with age and it is very sensible to related cognitive impairments.
- **Psychomobility:** It is the ability to coordinate the body movements in order to achieve behavioral objectives. Age-related physical and cognitive changes decrease this capability.

In summary, our studies aim to address the following research questions:

- What are short-term effects of attention training with elderly who are suffering from mild cognitive impairments?
- Will the cognitive stimulation game with the robot help the elderly individuals with mild cognitive impairments increase their "positive behavior" (e.g., smiling, speaking, and participating in group activities)?

The robots used in our system alternate verbal messages with non-verbal feedback in order to maintain the users' interest, use the name of the user for a more friendly and natural interaction, monitor users' performance, and provide real-time feedback as a function of the current and/or history performance so as to boost motivation for the execution of the exercise. The rest of this paper is structured as following: Section 2 depicts the first study for music-based cognitive stimulation; Section 3 illustrates the second ongoing study for cognitive attention stimulation; and finally Section 4 concludes the paper.

2 Study 1

This first study consists of the interaction between a robot and a user with dementia and/or Alzheimer's disease, with the main goal of helping users improve or maintain their cognitive attention through encouragements in a music-based cognitive stimulation game This approach consists of two parts: supervised learning and adaptation. The robot models the level of game challenge that can be: (a) Difficult: no hints; (b) Medium: when the song excerpt starts say "push the button" but do not indicate which button to push; and (c) Easy: when the song excerpt starts say which button to push. The supervised learning system learns the Accepted Variation Band (AVB) for each game level and for each disability bucket (mild, moderate, and severe), as a function of the user's task performance. The learning phase is followed by an adaptation phase, where the robot adapts its behavior so as to minimize the user's reaction time and maximize the correctness of the user's answers. If the user's task performance is below the Accepted Variation Band, the user is performing better than during the learning phase. The user is then promoted to the next level of game difficulty (if not already at the Difficult level). If the user's task performance is above the Accepted Variation Band, the user is not performing well enough. The user is then helped by having the game difficulty level decreased (if not already at the Easy level).

The Bandit robot from University of Southern California (USC) was used for the experiments.

The robot used arm movements to encourage the user in response to correct answers. The game challenge level, i.e., the amount of hints provided, was initialized based on the participant's level of impairment and adapted based on the participant's task performance (i.e., reaction time and number of correct answers).

The initial pilot experimental group consisted of 9 participants (4 male, 5 female), from our partner Silverado Senior Living care facility. All the participants were seniors over 70 years old suffering of cognitive impairment and/or Alzheimer's disease. The cognitive scores assessed by the SMMSE test were as follows: 1 mild, 1 moderate, and 7 severe. Due to the total unresponsiveness of 6 of the severely affected participants, only 1 severely cognitively disabled participant was retained for the rest of the study, resulting in a final participant group composed of 3 participants (all female). The results obtained over 6 months of robot interaction (excluding the 2 months of learning) suggest that the elderly people suffering of dementia and/or Alzheimer's can sustain attention to music across a long period of time (i.e., on average 20 minutes for mildly impaired participants, 14 minutes for moderately impaired participants, and 10 minutes for severely impaired participants) of listening activity designed for the dementia and/or Alzheimer's population.

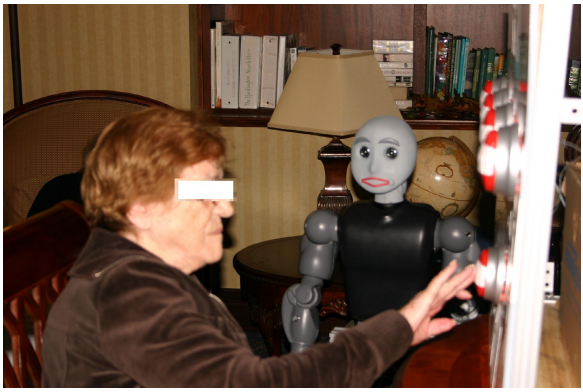


Fig. 1. Study 1: Participant playing the music game

Moreover, we found no adverse responses to the robot. The participants reported enjoying interacting with the robot and expressed anticipation of future sessions. At the end of the study, the three participants demonstrated attachment to the robot and the music game by referring to it as part of their weekly schedule and something they looked forward to. From the interviews with the participants' families and the facility staff, we can conclude that the robot had a positive effect on the users because they discussed it and described their interactions with it to both their families and other residents in the facility. The robot effectively became a part of the participants' narratives about their lives. Based on those discussions, we received inquiries from two other residents about participating in the music game with the robot. While it was not possible to include them in the study due to the lack of permission from their families, their interest was encouraging.

More about this study can be found in [16].

3 Study 2 - Ongoing Research

3.1 Experimental Design

Together with a group of psychologists and gerontologists from Carol Davila University of Medicine and Pharmacy (Bucharest, Romania) and Memory Clinic (Romania), we have designed various simple cognitive pairing games. The individuals with MCI have to classify different type of objects by following some simple rules. The scenario is the following: the user stands in front of a table where a number of objects are placed; he/she has to choose objects from the table by taking into consideration the rule of the game; the user takes an object and scans it with the barcode reader found on the table; once chosen, the object cannot be brought back again on the table; the game is finished when no objects are presented anymore on the table (however, the participants can stop the game at any time).

The social intelligent robot is present during the game and it monitors the user's task performance. The robot provides a feedback to the user as a function of his/her performance during the game, gives a global summary of the game, and finally suggests a new scenario that could be suitable for the individual in order to focus on the potential problems revealed by the current game.

The robot provides verbal and non-verbal feedback depending on the personality of the patient in order to stimulate him/her to continue and concentrate on the game. This is inspired by some of our previous studies based on the "similarity attraction principle" [12], [2] stating that individuals are attracted by others with the same personality traits.

There are different variations of the structure of the game: the rule is either established a priori by the application, the user has to conform to it and choose only those objects matching the rule, or the theme of the game is chosen by the user and the application establishes the rule for this configuration. Moreover, the game can be played with or without time limit. Some examples of rules are: combined-rule - "Select the objects that are small and gray.", or simple-rule - "Select the domestic animals.".

We built a database that contains important data about: each user (e.g., id, birth date, level of cognitive disability), the themes of the games chosen to be played (e.g., geometric shapes, animals), the attributes that describe each type of object (e.g., color, size, living environment), the actual objects with each specific barcode, the rules that can be applied in a game (e.g., "Choose all triangles"), the configuration of the play (theme, duration, feedback messages, objects), game results (i.e., a synthesis of a game trial). A configuration contains a general design of a game, which can be associated to each individual. We maintain the history of all game trials in order to be able to draw a conclusion about the exercise that is most suitable to be played in the next session. The database allows to conceive variations of the game on different themes according to the interests and the needs of the individuals (see Figure 2).

A real-time adaptive feedback module has been developed - the robot monitors the individual and based on the game rule provides a constructive feedback.

In this work, we designed three types of feedback: (1) positive feedback when the player's task perfromance is good, the robot encourages the individual to continue the good work; (2) negative feedback when the game performance decreases and/or many errors are made, the robot warns the player about his/her errors and encourages him/her to continue but to be more attentive (e.g., "Be careful, you have chosen wrong objects!"), and (3) a neutral feedback when the number of correct and wrong movements is similar or the delay between two movements is too long, the robot tries to stimulate the player. The different parameters refer to: the initial delay after the robot first's feedback, the delay between two feedbacks, and the match.

The dynamics of the different parameters is also given by the level of difficulty of the game: a game containing fewer objects is designed for a person with a higher level of cognitive impairment, in which case a feedback at short intervals is provided in order to guide him/her to reach the solution. Whilst, a game with more objects to classify is created for persons with a mild impairment that can handle the play easier and do not need to be helped so frequently.

The task performance is a summary of the entire game, which brings together the information captured by the system: the objects chosen, the distribution into three categories: correct, wrong, and omitted objects, the number of each of these categories, the total time of the game, and the average response time. Each user has his/her own history. This can provide an idea about the evolution of the individual's task performance over time. This history is an indication of which exercises were more useful and an intuition about which is the path that should be followed in order to maintain user's cognitive attention.

3.2 Game Adaptation

Based on the user's history the system can see the type of game in which an individual has more difficulties and therefore decide which configuration in the database is the most suitable for the user to execute next.

Our database contains tagged game trials with the most important attributes that describe them for each individual. These can be used to decide, which of the test is more suitable for a certain individual. The game trials are categorized into three levels: simple, advanced, and complex. These levels are established mainly by taking into consideration the number of objects in the configuration and the complexity of the rule. The tags that a configuration can have are the following: general, simple rule, combined rule, speed, and contrast. They have the following significance: general - the exercise is a general one, based on it we can identify a more specific problem of the individual; simple rule the configuration requires to comply with a simple rule; combined rule the rule imposes some restrictions on more than one attribute; speed the most important aspect of the exercise is its maximum duration, and it evaluates the speed of the user in deciding which are the correct objects; and contrast the set of objects is composed mainly of objects of two types because the user might have difficulties in distinguishing between two values of an attribute (e.g. he/she confuses red and orange).

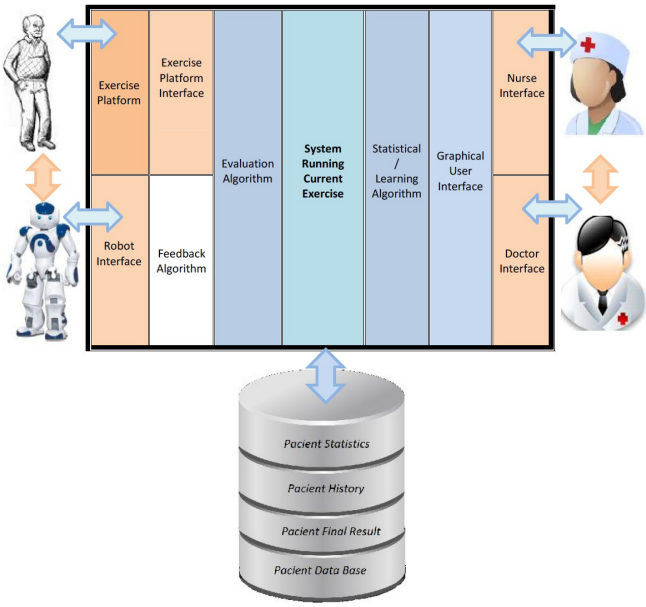


Fig. 2. Study 2: System Architecture

Based on the most recent result of a game trial and helped by the history results, the system decides the game that the individual has problem solving and suggests a new exercise that should help him/her to overcome it by practicing.

3.3 System Testbed

Our system is composed of several components:

Barcode Reader. A barcode reader (IBC-2000-N-USB) is used to scan the various objects used in the games. The objects are identified in the database by their attributes, number, and a barcode. This code contains information about the characteristics of the object represented by the identifiers in the database: the class of the object, the type and the values of the attributes. The user scans each chosen object and the robot receives the sequence of barcodes. Based on this information the robot is capable of understanding and evaluating the task performance of the user and provide an appropriate feedback.

Robot Testbed. The experimental test-bed used in this study is the humanoid Nao robot developed by Aldebaran Robotics[1]. Nao is a 25 degrees of freedom robot, equipped with an inertial sensor, two cameras, eyes eight full-color RGB

[1] `http://www.aldebaran-robotics.com/`

LEDs, and many other sensors, including a sonar which allows it to comprehend its environment with stability and precision. Nao is also equipped with a voice synthetizer and 2 speakers enabling vocal communication.

This is an ongoing study. A group of 6 patients from the Memory Clinic (Bucharest, Romania) was selected for the experiments. More results will be reported for this study at a later stage.

4 Conclusions

The presented research presented some studies that we have developed for providing customized cognitive stimulation for elderly users with Alzheimer's disease and or mild cognitive impairment (MCI). Our work involved the use of a robotic system in the context of different cognitive games. The first pilot study demonstrated the overall effect of sustaining or improving user performance on the memory task around a music-based cognitive game through the assistance of a robotic system over the period of up to 8 months. The results are encouraging in light of our pursuit toward creating personalized socially assistive technologies that aim to improve human quality of life. The second study is an ongoing work and the system will soon be tested with a group of individuals from the Memory Clinic (Bucharest, Romania).

References

1. Albert, M.S., Dekosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P., Carrillo, M.C., Thies, M., Phelps, C.H.: The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup (April 16, 2011) (Article in Press)
2. Aly, A., Tapus, A.: A Model for Synthesizing a Combined Verbal and NonVerbal Behavior Based on Personality Traits in Human-Robot Interaction. In: Proceedings of the ACM/IEEE Human-Robot Interaction Conference (HRI 2013), Tokyo, Japan (March 2013)
3. Clair, A., Ebberts, A.: The effects of music therapy on interactions between family caregivers and their care receivers with late stage dementia. Journal of Music Therapy 34, 148–164 (1997)
4. Eysenck, H.J.: The structure of human personality (1953)
5. Fasola, J., Matarić, M.J.: Using Socially Assistive HumanRobot Interaction to Motivate Physical Exercise for Older Adults. Proceedings of the IEEE 100(8) (August 2012)
6. Feil-Seifer, D., Matarić, M.J.: Defining socially assistive robotics. In: Proc. IEEE International Conference on Rehabilitation Robotics (ICORR 2005), Chicago, Il, USA, pp. 465–468 (June 2005)
7. Granata, C., Chetouani, M., Tapus, A., Bidaud, P., Dupourque, V.: Voice and graphical -based interfaces for interaction with a robot dedicated to elderly and people with cognitive disorders. In: Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Viareggio, Italy (September-October 2010)

8. Heerink, M., Krse, B., Evers, V., Wielinga, B.: Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model. International Journal of Social Robotics 2, 361–375 (2010)
9. Kidd, C., Taggart, W., Turkle, S.: A sociable robot to encourage social interaction among the elderly. In: IEEE International Conference on Robotics and Automation (ICRA), Orlando, USA (May 2006)
10. Libin, A., Cohen-Mansfield, J.: Therapeutic robocat for nursing home residents with dementia: Preliminary inquiry. Am. J. Alzheimers Dis. Other Demen. 19(2), 111–116 (2004)
11. Marti, P., Giusti, L., Bacigalupo, M.: Dialogues beyond words. Interaction Studies (2008)
12. Isbister, K., Nass, C.: Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. International Journal of Human Computer Studies 53(2), 251–268 (2000)
13. Can alzheimer's disease be prevented? U.S. National Institute of Health. National Institute of Aging (2008)
14. Tapus, A., Tapus, C., Matarić, M.J.: User-Robot Personality Matching and Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy. In: Intelligent Service Robotics, Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics, vol. 1(2), pp. 169–183 (2008)
15. Tapus, A., Tapus, C., Matarić, M.: The Use of Socially Assistive Robots in the Design of Intelligent Cognitive Therapies for People with Dementia. In: Proceedings of the International Conference on Rehabilitation Robotics (ICORR), Kyoto, Japan (June 2009)
16. Tapus, A., Tapus, C., Matarić, M.: The Role of Physical Embodiment of a Therapist Robot for Individuals with Cognitive Impairments. In: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Toyama, Japan (September-October 2009)

The Intelligent Room for Elderly Care

Oscar Martinez Mozos[1], Tokuo Tsuji[2], Hyunuk Chae[2], Shunya Kuwahata[2], YoonSeok Pyo[2], Tsutomu Hasegawa[2], Ken'ichi Morooka[2], and Ryo Kurazume[2]

[1] School of Computer Science, University of Lincoln, Lincoln LN6 7TS, United Kingdom
[2] Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan
omozos@irvs.ait.kyushu-u.ac.jp
http://fortune.ait.kyushu-u.ac.jp

Abstract. Daily life assistance for elderly is one of the most promising and interesting scenarios for advanced technologies in the near future. Improving the quality of life of elderly is also some of the first priorities in modern countries and societies where the percentage of elder people is rapidly increasing due mainly to great improvements in medicine during the last decades. In this paper, we present an overview of our informationally structured room that supports daily life activities of elderly with the aim of improving their quality of life. Our environment contains different distributed sensors including a floor sensing system and several intelligent cabinets. Sensor information is sent to a centralized management system which processes the data and makes it available to a service robot which assists the people in the room. One important restriction in our intelligent environment is to maintain a small number of sensors to avoid interfering with the daily activities of people and to reduce as much as possible the invasion of their privacy. In addition we discuss some experiments using our real environment and robot.

Keywords: Quality of Life Technologies, Assistive Robotics, Intelligent Room, Ambient Intelligence.

1 Introduction

Inside the many applications related to quality of life technologies, elderly care is one of the most promising ones both in social and economic terms. Improving the quality of life of elderly is also some of the first priorities in modern countries and societies where the percentage of elder people is rapidly increasing due mainly to great improvements in medicine during the last decades.

Different intelligent systems are being developed to assist elderly in their daily life environment. The main idea of these scenarios is to gather information about the environment of the people so that an intelligent system can keep track of their actions and their surroundings and can act when the person needs some assistance. The help received by the person can be on demand or alternatively the system can decide by itself when to take an action in order to assist people [16,7,4,8,13,14,11]. In addition, service robots can be available to assist people alongside the intelligent

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 103–112, 2013.

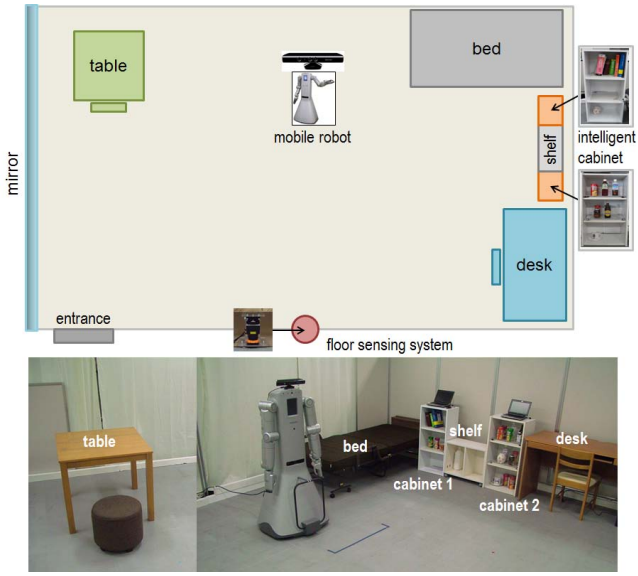


Fig. 1. The top image outlines a map of our intelligent room. Real images of the different components are show in the bottom images.

environment. Actually, it is expected that service robots will soon be playing a role of companion to elderly people, or a role of assistant to humans with special needs at home [3], [17], [6], [2], [9], [5]. In particular, one of the most demanding tasks by users will be the go-and-fetch of objects that are needed for their everyday activities [18], [1].

This paper presents an overview of our informationally structured room which aims to assist elder people in their daily life. Our environment contains different distributed sensors including a floor sensing system and several intelligent cabinets as shown in Fig.1 Sensor information is sent to a centralized management system which processes the data and makes it available to a service robot which assists the people in the room. One important restriction in our intelligent environment is to maintain a small number of sensors to avoid interfering with the daily activity of people and to reduce as much as possible the invasion of their privacy. For this reason we restrict the use of the camera on the robot to only some predetermined situations.

2 The Informationally Structured Room

This section briefly describes the different components of our informationally structured environment. In particular, our scenario represents a room in a house as shown in Fig. 1. The room contains two intelligent cabinets, one shelf, a bed, a desk with a chair, and a dining table. In addition, the room is equipped with a floor sensing system for object and people detection.

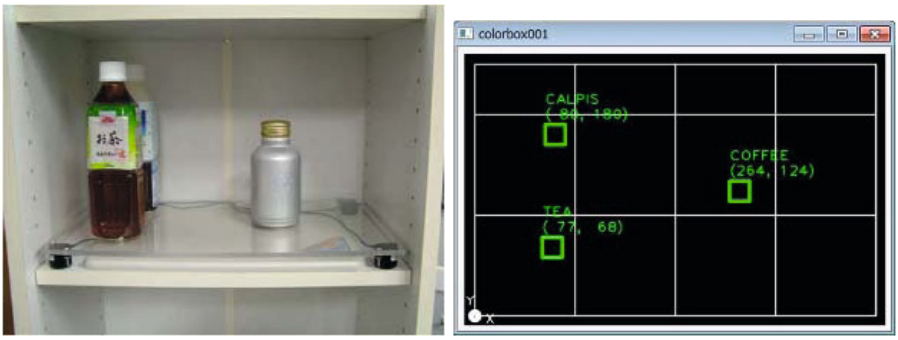


Fig. 2. Information about objects provided by the intelligent cabinet. Squares in the screen indicate the position of the different objects together with their description.

2.1 Intelligent Cabinets

The cabinets installed in our room (see Fig.1) are equipped with RFID readers and load cells to detect the type and position of objects inside. Every object in our environment is attached one RFID tag containing a unique ID that identifies the object. This ID is used to retrieve the attributes of the object in our database. Using the RFID readers we can detect the presence of a new object inside the cabinet. In addition, the information of the load cells allows us to determine its exact position inside the cabinet. An example detection of objects in one intelligent cabinet is shown in Fig. 2. Further details about our intelligent cabinets can be found in [10].

2.2 Floor Sensing System

In addition to the intelligent cabinets our room is equipped with a floor sensing system used to detect objects on the floor and people walking around. This sensing systems is composed of a laser range finder which is located on one side of the room as shown in Fig.1. Moreover, a mirror is installed along one side of the room to help the detection when clutter occurs. This configuration allows a reduction of dead angles of the LRF and it is more robust against occlusions [15]. An example detection of an object using this system is shown in Fig. 3.

People tracking is performed by first applying static background subtraction and then extracting blobs in the rest of the measurements. Blobs are later tracked by applying a Kalman filter by matching profiles of blobs corresponding to legs, and extending the motion using accelerations of legs [10].

2.3 Town Management System

The previous sensing system and the robot itself are connected to our *Town Management System* (TMS), which integrates sensor data into an online environment

Fig. 3. The left image depicts an example of an object detected on the floor by our floor sensing system. The right image shows the detection of a person indicated by a white square.

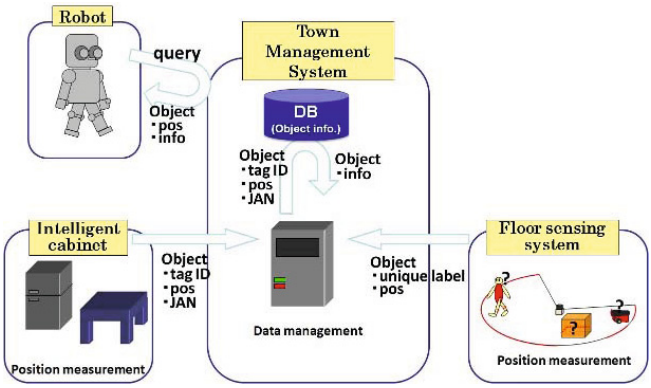


Fig. 4. Information flow between the TMS and the sensors and robot in our informationally strucutred room

database and provides robots with real-time information on its dynamically changing surroundings. The TMS was originally designed to include informationa about distributed sensors and robots in a wider environment [12]. This central database management system provides information about indoor maps, RFID tag ids and related information, and notification of predefined environment events and their occurrence. The information flow between our intelligent room and the TMS is shown in Fig. 4.

3 Service Robot

Finally, the person acting in the room is assisted by a SmartPal humanoid robot (Fig. 5) from Yaskawa Electric Corporation. This robot will be responsible for fetching objects or pointing to them. The robot is composed of a mobile platform, two arms with seven joints, and one-joint grippers used as hands. In addition we

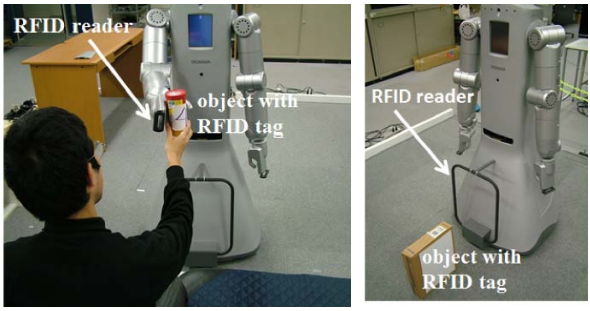


Fig. 5. Assistive humanoid robot SmartPal equipped with RFID readers

equipped the robot with a RGB-D camera which is used for object recognition in restricted regions of interest, and only under specific requests. In order to maintain the privacy of people we do not use this camera for general vision purposes. Additional RFID readers are situated on the hands and front of the robot (Fig. 5) to recognize grasped objects and objects on the floor.

3.1 Visual Memory for Object Searching

Our service robot is equipped with a visual memory system which helps in the task of object searching and finding. The visual memory system is used by the robot to detect changes on predefined places where objects usually appear. In our case we restrict the application of this visual system to the table of our intelligent room (see Fig. 1). The reason for that is to keep the privacy of the people as much as possible and to avoid registering images of the user during his daily and private activities.

The visual memory system is composed of two main steps. In the first one changes are detected in the area of interest which usually correspond to appearance, disappearance or movement of objects. In a second step the areas corresponding to the changes are analyzed and new objects are categorized. The complete visual memory system is shown in Fig. 6.

3.2 Change Detection

The first step of our visual memory is responsible for detecting changes in the area of interest, which is a table in our case. The change detector works as follows. At some point in time t_1 the service robot takes a snapshot z_1 of a table. Since we use a kinect camera then our observation is composed of a 3D point cloud. At some later point in time t_2 the robot takes a second snapshot z_2 of the same table. The positions p_1 and p_2 of the robot during each observation are known and determined our localization system so that we can situated each observation in a global reference system. In addition, we improve the alignment of the two point clouds using the ICP algorithm. This step allows us to correct small errors that can occur in our localization system.

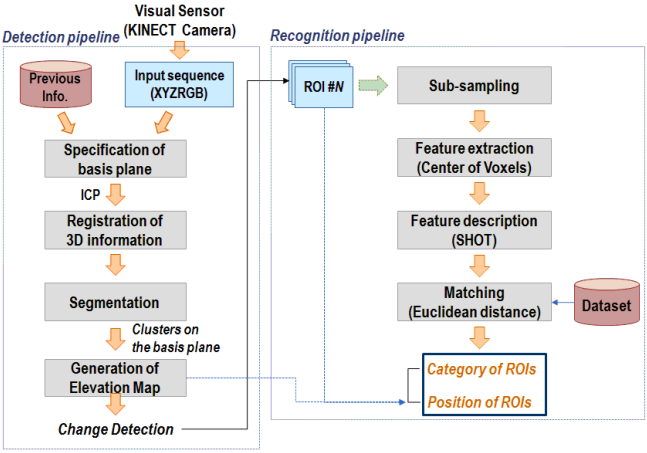


Fig. 6. Schematic diagram for change detection and its categorization

For each independent view, we extract the plane corresponding to the top of the table by applying a method based on RANSAC. The remaining points, which pertain to the possible objects on top of the table, are projected to a 2D grid. The cells in this grid are clustered using connected components and each resulting cluster is assumed to be the 2D representation of a different object on the table. We then compare the 2D clusters in each view and determine the different clusters between the two views which correspond to changes on the table. A resulting change detection is shown in Fig. 7

3.3 Object Categorization

The point clusters corresponding to possible changes on the table are further categorized into a set of predefined set of object categories contained in our database as shown in Fig. 8. Our method finds the best matching between the cluster representing a change and the cluster representing each object in our dataset. Our 3D matching method is based on correspondence grouping [20] using the SHOT 3D surface descriptor [19] as key point descriptor. The best matching is obtained as the minimum distance according to

$$D = \frac{corr}{max(N_{model_j}, N_{cluster})}, \tag{1}$$

where *corr* represent the number of correspondences between keypoints in the model of our dataset and the keypoints in the cluster, N_{model_j} indicates the number of keypoints found in the model, and N_{model} represents the number of keypoints found in the cluster. Figure 9 shows an example result of the complete process of change detection and categorization.

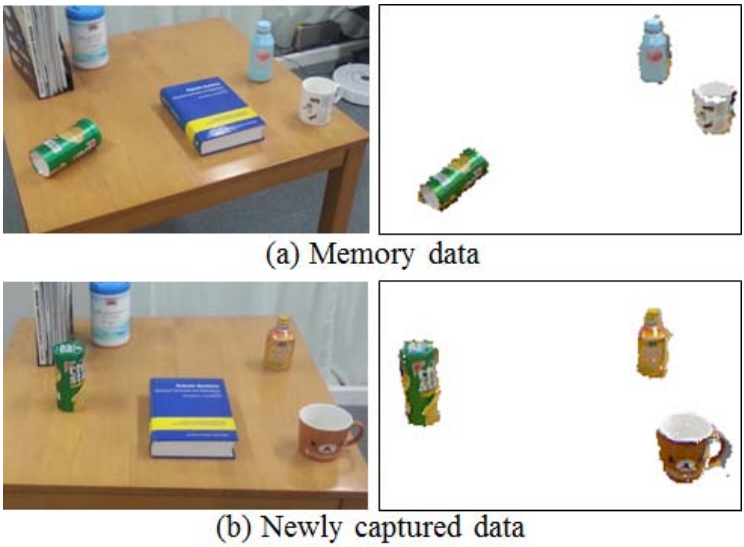


Fig. 7. Changes detected between two consecutive views of a table

3.4 Grasp Planning

Once an object is found in the environment the robot usually needs to grasp it to deliver it to the user. In our system we approximate objects using a polygon model. Using this model the robot should find a feasible grasp posture automatically and quickly. Therefore we use two approximation boxes for both the object and the robot hand. The first box is composed of the bounding box called *object convex polygon* (OCP) which includes the object in object coordinate system. For large object, we split the object into several regions and calculate the OCP for each region. The second box which means the capacity of object size for the hand is also defined in the hand coordinate system as a *grasp rectangular convex* (GRC). Using these boxes our planner determines the hand position/orientation by checking whether the position/orientation of GRC can include the OCP. For selecting one grasping posture from multiple candidates of the position/orientation of the hand, we evaluate the movement capability of inverse kinematics solution.

4 Discusion and Future Work

In this paper we have introduced our informationally structured room which is designed to support daily activities of elder people. The room contains several sensors to monitor the environment and the person. Moreover, the person is assisted by a humanoid robot which uses the information of the environment to support different activities. In addition, we want to stress the importance of keeping the privacy of people during their daily activities and the need to reduce the invasion of their privacy as much as possible.

Fig. 8. Dataset of daily life objects for our environment

Fig. 9. Resulting matching of changes on a table. The top row indicates the label of the changed objects. The bottom row shows the categorization using our method. The pet bottle is misclassified as a chip container.

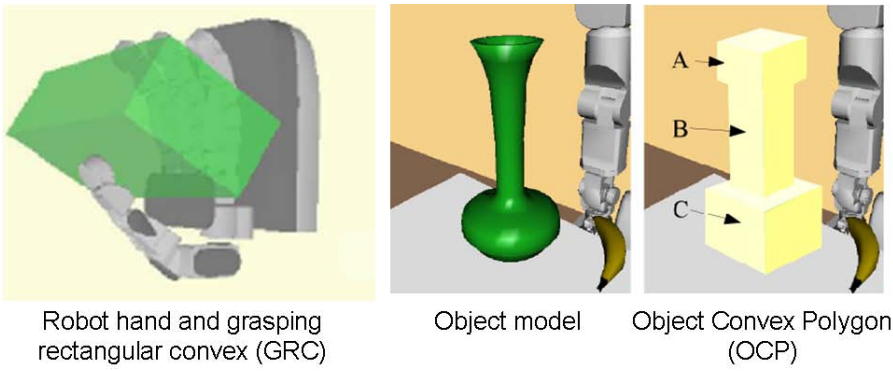


Fig. 10. Aa grasp rectangular convex for a hand, an object model of a base, and object convex polygons of the model

In this work we have concentrated on the go-and-fetch task which we prognosticate to be one of the most demanding tasks by elderly in their daily life. In this respect we have presented the different subsystems that are implicated in this task, and have showed several independent short-term experiments to demonstrate the suitability of the different subsystems. In the future we aim to design and prepare a long-term experiment in which we can test the complete system for a longer period of time.

References

1. Dario, P., Guglielmelli, E., Laschi, C., Teti, G.: Movaid: A personal robot in everyday life of disabled and elderly people. Technology and Disability 10, 77–93 (1999)
2. Hasegawa, T., Muarkami, K.: Robot town project: Supporting robots in an environment with its structured information. In: Proc. of the 5th Int. Conf. on Ubiquitous Robots and Ambient Intelligence, pp. 119–123 (2006)
3. Kawamura, K., Iskarous, M.: Trends in service robots for the disabled and the elderly. In: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 3, pp. 1647–1654 (September 1994)
4. Kayama, K., Yairi, I.E., Igi, S.: Semi-autonomous outdoor mobility support system for elderly and disabled people. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2606–2611 (2003)
5. Kim, B.K., Tomokuni, N., Ohara, K., Tanikawa, T., Ohba, K., Hirai, S.: Ubiquitous localization and mapping for robots with ambient intelligence. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 4809–4814 (2006)
6. Kim, M., Kim, S., Park, S., Choi, M.-T., Kim, M., Gomaa, H.: Service robot for the elderly. IEEE Robotics Automation Magazine 16(1), 34–45 (2009)
7. Lee, J.H., Hashimoto, H.: Intelligent space - concept and contents. Advanced Robotics 16(3), 265–280 (2002)
8. Mori, T., Takada, A., Noguchi, H., Harada, T., Sato, T.: Behavior prediction based on daily-life record database in distributed sensing space. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 1833–1839 (2005)
9. Murakami, K., Hasegawa, T., Kurazume, R., Kimuro, Y.: A structured environment with sensor networks for intelligent robots. In: Proc. of IEEE Int. Conf. on Sensors, pp. 705–708 (2008)
10. Murakami, K., Hasegawa, T., Shigematsu, K., Sueyasu, F., Nohara, Y., Ahn, B.W., Kurazume, R.: Position tracking system of everyday objects in an everyday environment. In: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3712–3718 (October 2010)
11. Murakami, K., Matsuo, K., Hasegawa, T., Kurazume, R.: Position tracking and recognition of everyday objects by using sensors embedded in an environment and mounted on mobile robots. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp. 2210–2216 (May 2012)
12. Murakami, K., Hasegawa, T., Kurazume, R., Kimuro, Y.: Supporting robotic activities in informationally structured environment with distributed sensors and rfid tags. Journal of Robotics and Mechatronics 21(4), 453–459 (2009)
13. Nakauchi, Y., Noguchi, K., Somwong, P., Matsubara, T., Namatame, A.: Vivid room: Human intention detection and activity support environment for ubiquitous autonomy. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 773–778 (2003)

14. Nishida, Y., Aizawa, H., Hori, T., Hoffman, N.H., Kanade, T., Kakikura, M.: 3d ultrasonic tagging system for observing human activity. In: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 1, pp. 785–791 (October 2003)
15. Nohara, Y., Hasegawa, T., Murakami, K.: Floor sensing system using laser range finder and mirror for localizing daily life commodities. In: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1030–1035 (October 2010)
16. Nugent, C.D., Finlay, D.D., Fiorini, P., Tsumaki, Y., Prassler, E.: Home automation as a means of independent living. IEEE Trans. Automation Science and Engineering 5(1), 1–9 (2008)
17. Roy, N., Baltus, G., Fox, D., Gemperle, F., Goetz, J., Hirsch, T., Magaritis, D., Montemerlo, M., Pineau, J., Schulte, J., Thrun, S.: Towards personal service robots for the elderly. Carnegie Mellon University (2000)
18. Srinivasa, S., Ferguson, D., Helfrich, C., Berenson, D., Collet, A., Diankov, R., Gallagher, G., Hollinger, G., Kuffner, J., VandeWeghe, M.: Herb: A home exploring robotic butler. Autonomous Robots 28, 5–20 (2010)
19. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 356–369. Springer, Heidelberg (2010)
20. Tombari, F., Di Stefano, L.: Object recognition in 3d scenes with occlusions and clutter by hough voting. In: 4th Pacific-Rim Symposium on Image and Video Technology (2010)

A Method for Estimating Patient Specific Parameters for Simulation of Tissue Deformation by Finite Element Analysis

Ken'ichi Morooka, Shuji Sonoki, Ryo Kurazume, and Tsutomu Hasegawa

Kyushu University, Motooka 744, Nishi-ku, Fukuoka 819-0395, Japan
morooka@ait.kyushu-u.ac.jp
http://fortune.ait.kyushu-u.ac.jp/~morooka/

Abstract. This paper proposes a method for estimating patient-specific material parameters used in the finite element analysis which simulates soft tissue deformation. The estimation of suitable material parameters for a patient is important for a navigation system for endoscopic surgery. At first, many data of soft tissue deformation are generated by changing the material parameters. Next, using Principle Component Analysis, each data with high dimensional is converted into the lower vector. The relationship between the material parameter and the deformation is found in the lower potential space.

Keywords: FE analysis, patient specific parameter, minimally invasive surgery, surgical navigation.

1 Introduction

Laparoscopic surgery is one of minimally invasive surgeries in which operates in abdomen. Since laparoscopic surgery is smaller invasive than traditional open surgery, recovery time and return to normal activities is shorter for patients. Therefore, laparoscopic surgery contributes to the improvement of patients' quality of life. On the contrary, performing laparoscopic procedures require special surgical skills. For example, unlike traditional open surgery, laparoscopic surgeons can not see and touch the patient's tissues directly. The surgeons estimate the 3D structure of the patient by only moving 2D laparoscope. In addition, the laparoscope has a small field of view. The lack of depth perception and the small visual field may prevent the surgeons from accessing the target tissue or the tumor without damaging neighbor tissues.

Recently, navigation systems for minimally invasive surgery have been developed [2,5,6,12]. One of components of the system is to display virtual images by superimposing the models into real endoscopic images. The models contain the internal information of the patient body such as the tumors within the tissue and/or the blood vessels behind the tissue. Such information can never be obtained from the endoscopic images. The use of the virtual images enables surgeons to approach the tissues or tumors safely and accurately.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 113–120, 2013.

During the surgery, the laparoscopic surgeons deal with soft tissues by special instruments. When a contact occurs between surgical instruments and tissues, the tissues are deformed according to their biomechanical behavior. To provide suitable virtual images, the navigation system needs to simulate the behavior of the tissue. The finite element method (FEM) is a well-known technique for accurately modeling the behaviors of continuous objects. Moreover, we have been developing a technique for real-time finite element analysis by multiple neural networks [8,9]. The formulation of FE analysis includes several parameters which define the behavior of the target object. In the case of the soft tissues, the setting of the parameters depends patient-related attributes including disease, age, gender, and so on. The determination of suitable values for the parameters is important for the system to provide the reliable tissue deformation. However, there are few methods [3,4] for estimating the appropriate parameter values during the operation. In the conventional navigation system, the parameters are given by a priori data from other measurements.

This paper proposes a new method for determining a patient-specific parameter values by observing the tissue deformations. In this paper, liver is used as a target soft tissue. At first, we model the relationship between the FE parameters and the motion sequence of the liver calculated by the FE analysis. When the motion of the liver with unknown is observed, the suitable parameters are estimated by using the model.

2 Basic Notation

2.1 Finite Element Parameters

Nonlinear FEM achieves a more physically realistic simulation for deforming continuous object with nonlinear material properties. To apply the nonlinear FEM to a target soft tissue, a constitutive model is needed to describe the mechanical properties of the tissue. Most materials can be regarded as incompressible hyperelastic solids [1]. An incompressible material preserves its volume during the deformation. A material is hyperelastic if there exists a potential function W such that a stress $\boldsymbol{S}$ is formulated by the derivative of W with respect to a strain $\boldsymbol{G}$:

$$\boldsymbol{S} = \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{G}}. \tag{1}$$

In our method, a Mooney-Rivlin material model was employed as the potential function:

$$W = c_{10}(J_1 - 3) + c_{01}(J_2 - 3) + c_{20}(J_1 - 3)^2 + c_{11}(J_1 - 3)(J_2 - 3) + c_{02}(J_2 - 3)^3, \tag{2}$$

where J_1 and J_2 are the first and second invariants of the right Cauchy-Green deformation tensor. In the case of a liver, the coefficients c_{10}, c_{01}, c_{11}, c_{20} and c_{02} are set to $c_{10} = 7.51 \times 10^4$, $c_{20} = 3.43 \times 10^5$, and $c_{01} = c_{11} = c_{02} = 0$ [7,11]. Therefore, $\boldsymbol{c} = [c_{10}, c_{20}]^T$ are regarded as the FE parameters in our method.

2.2 Deformation Pattern

When a specific force acts on the liver, the FE parameters are estimated by matching the real behavior of the liver and the liver deformation patterns generated beforehand. Each deformation pattern is obtained by applying the nonlinear FE analysis to the liver model with material parameter values when the specific force is given. In our analysis, the liver deformation is simulated when the left lobe of the liver is lifted by the forceps. The output of the analysis is the sequence of the deformed models from start $f = 1$ to end frame $f = F$.

When the liver model contains N_v nodes on the model surface, the sequence of the deformed models is represented by two kinds of vectors. Given two successive deformed models $\mathcal{M}^{(f)}$ and $\mathcal{M}^{(f+1)}$ $(f = 1 \cdots F-1)$, the first vector is a displacement vector $\boldsymbol{D}^{(f)}$ obtained by collecting the displacement of the node on the surface of the models:

$$\boldsymbol{D}^{(f)} = \left[{\boldsymbol{d}_1^{(f)}}^T, \cdots, {\boldsymbol{d}_{N_v}^{(f)}}^T \right]^T; \tag{3}$$

where $\boldsymbol{d}_j = (x_j, y_j, z_j)^T$ is the displacement of the j-th $(j = 1, \cdots, N_v)$ node on the surface of the model $\mathcal{M}$. A vector (or a matrix) $\boldsymbol{A}^T$ means the transpose of $\boldsymbol{A}$. The second vector is the velocity vector $\boldsymbol{V}^{(f)}$ defined by

$$\boldsymbol{V}^{(f)} = \boldsymbol{D}^{(f+1)} - \boldsymbol{D}^{(f)} = \left[[\boldsymbol{d}_1^{(f+1)} - \boldsymbol{d}_1^{(f)}]^T, \cdots, [\boldsymbol{d}_{N_v}^{(f+1)} - \boldsymbol{d}_{N_v}^{(f)}]^T \right]^T. \tag{4}$$

A deformation vector $\boldsymbol{m}^{(f)}$ is the set of the two vectors;

$$\boldsymbol{m}^{(f)} = \left[(\boldsymbol{D}^{(f)})^T, (\boldsymbol{V}^{(f)})^T \right]^T \tag{5}$$

Using eq.(5), the sequential deformation data is described as the $(F-1) \times 2N_v$ dimensional matrix by collecting all the deformation vectors as follows:

$$\boldsymbol{M} = \left[\boldsymbol{m}^{(1)}, \cdots, \boldsymbol{m}^{(F-1)} \right]^T \tag{6}$$

All the element of the displacement and velocity vectors are normalized by the maximum magnitude d_{max} and v_{max} of the displacement and the velocity vectors, respectively.

K deformation patterns are generated by the simulation using K kinds of the FE parameter values. For each simulation, the parameter values are randomly selected from the range $[\bar{c} \pm 0.25\bar{c}]$ [10]. Here, the vector $\bar{c}$ means the FE parameter value of a normal liver [11]. Fig. 1 shows the examples of the deformations of the liver model with different material parameters.

3 Patient-Specific Parameter Estimation

Since the dimension of each deformation vector is high, the deformation vector is mapped onto a lower dimensional space to deal with the vector easily. The

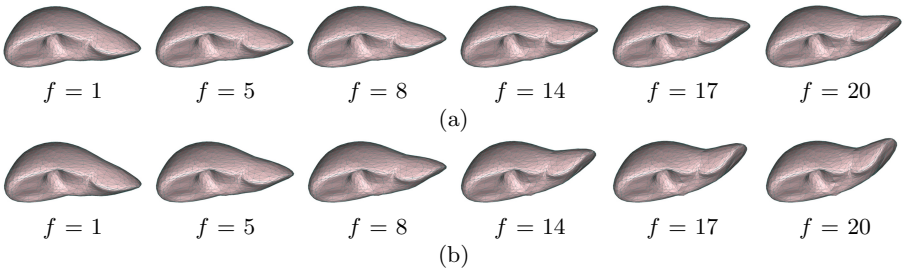


Fig. 1. Examples of the deformation patterns of a liver with different kinds of the FE parameters: (a) $c_{10} = 5.63 \times 10^4$, $c_{20} = 2.57 \times 10^4$; (b) $c_{10} = 9.38 \times 10^4$, $c_{20} = 4.29 \times 10^4$

lower dimensional space, called the potential space, is constructed by using the deformation patterns of the normal liver. Practically,

Practically, when there are the K' deformation patterns $\boldsymbol{M}'_i$ $(i = 1, \cdots, K')$ of the normal liver, a matrix $\boldsymbol{X}$ is computed by

$$\boldsymbol{X} = \boldsymbol{Y}^T \boldsymbol{Y}; \tag{7}$$
$$\boldsymbol{Y} = [\boldsymbol{M}'_1 - \bar{\boldsymbol{M}},\ \boldsymbol{M}'_2 - \bar{\boldsymbol{M}}, \cdots\ \boldsymbol{M}'_{K'} - \bar{\boldsymbol{M}}].$$

$$\bar{\boldsymbol{M}} = \frac{1}{K'} \sum_{i=1}^{K'} \boldsymbol{M}'_i. \tag{8}$$

The eigenvectors of $\boldsymbol{X}$ are calculated by applying Principle Component Analysis (PCA) to $\boldsymbol{X}$. Each eigenvector has its eigenvalue which means the distribution of the deformation patterns along the eigenvector. By selecting a compact set of the eigenvectors, the deformation patterns can be translated into a lower-dimensionality subspace within which most of the deformation patterns can be represented with acceptable accuracy. To find this, we sort all the eigenvectors into descending order of their eigenvalues. The first two eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are selected, and forms the potential space. Using the selected eigenvectors, $6N_v = (2 \times 3N_v)$ projection matrix $\boldsymbol{\Phi}$ to the potential space is defined as

$$\boldsymbol{\Phi} = [\boldsymbol{e}_1\ \boldsymbol{e}_2]^T. \tag{9}$$

Using the matrix $\boldsymbol{\Phi}$, an arbitrary deformation patterns $\boldsymbol{M}$ with $6N_v$ dimensional is converted into a lower 2-dimensional vector $\boldsymbol{M}^*$ in the potential space:

$$\boldsymbol{M}^* = \left[p_1\ p_2\right]^T = \boldsymbol{\Phi}(\boldsymbol{M} - \bar{\boldsymbol{M}}). \tag{10}$$

Fig. 2 shows two examples of the trajectories of the deformation patterns.

The trajectory of the deformation patterns is described as parametric line functions. In our method, the deformation patterns in the potential space is formulated by a four dimensional polynomial curve function:

$$p_2 = u_1 p_1^4 + u_2 p_1^3 + u_3 p_1^2 + u_4 p_1 + u_5 = \boldsymbol{u}^T \boldsymbol{p}_1 \tag{11}$$

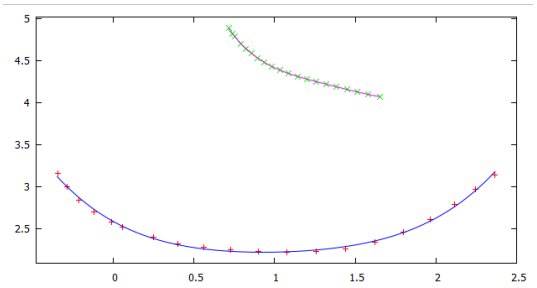


Fig. 2. Examples of the deformation patterns mapped onto the potential space

where $\boldsymbol{p}_1 = [p_1^4,\ p_1^3,\ p_1^2,\ p_1,\ 1]^T$. The vector $\boldsymbol{u} = [u_1,\ u_2,\ u_3,\ u_4,\ u_5]^T$ is the set of coefficients of the curve function. After the vector $\boldsymbol{u}$ is obtained, the FE parameter set $\boldsymbol{c} = [c_{10}, c_{20}]^T$ is estimated by

$$\boldsymbol{c} = \begin{bmatrix} c_{10} \\ c_{20} \end{bmatrix} = \begin{bmatrix} \lambda_{11}u_1 + \lambda_{12}u_2 + \lambda_{13}u_3 + \lambda_{14}u_4 + \lambda_{15}u_5 + \lambda_{16} \\ \lambda_{21}u_1 + \lambda_{22}u_2 + \lambda_{23}u_3 + \lambda_{24}u_4 + \lambda_{25}u_5 + \lambda_{26} \end{bmatrix} = \boldsymbol{\Lambda} \begin{bmatrix} \boldsymbol{u} \\ 1 \end{bmatrix}. \quad (12)$$

Therefore, the matrix $\boldsymbol{\Lambda}$ is calculated by using many deformation patterns with different FE parameters and eqs (11) and (12).

When the deformation of a liver with unknown FE parameters is observed, all the deformation vectors are mapped onto the potential space by using eq.(10). A least-square technique finds the coefficients $\boldsymbol{u}$ so that the curve function fits the trajectory in the potential space. The FE parameters are estimated by multiplying the obtained coefficients $\boldsymbol{u}$ with the matrix $\boldsymbol{\Lambda}$ in eq.(12).

4 Experimental Result

To verify the applicability of the proposed method, we made some experiments using a liver model composed of 4,804 nodes and 15,616 tetrahedral elements. The models are created by the commercially available softwares (CDAJ-Modeler CFD, CD-adapco JAPAN Co., LTD.) while their deformation patterns are generated by the FEM analysis software ("Marc" produced by MSC.Software Co.). Each deformation pattern consists of 21 deformed models along time. 81 deformation patterns are generated by setting the FE parameters to $c_{10} = 7.51 \times 10^4$ and $c_{20} = 3.43 \times 10^4$, and selecting randomly from from the range $[\boldsymbol{c} \pm 0.25\boldsymbol{c}]$.

The first experiment is to estimate unknown FE parameters which are not used in the construction of the projection matrix $\boldsymbol{\Lambda}$ in eq.(12). In the experiment, 16 test deformation patters are generated. Table. 1 (a) shows the three results of the estimated parameters by the proposed method. Our estimated values of the parameters is the almost same as the true values. The deformations are simulated by using the true and estimated parameter values, and the obtained models, called our models, are compared with the standard model generated by

Table 1. Estimation of FE parameters by our method with unknown deformation data with (a) no noise and (b) much noise

Test Pattern	True value [10^4]		Estimated value [10^4]	
	c_{10}	c_{20}	c_{10}	c_{20}
A	7.08	3.41	7.07	3.41
B	8.43	4.02	8.43	4.01
C	5.71	7.91	5.99	8.07

(a)

Data Pattern	True value [10^4]		Estimated value [10^4]	
	c_{10}	c_{20}	c_{10}	c_{20}
A	7.85	2.65	7.88	2.42
B	5.71	7.91	5.77	7.83
C	9.34	2.99	9.67	0.73

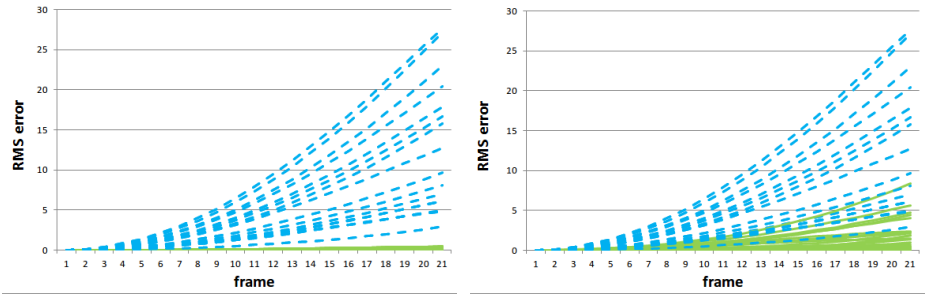
(b)

the simulation using the true values. Here, the difference E between two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are formulated by

$$E(\mathcal{M}_1, \mathcal{M}_2) = \frac{1}{FN_v} \sum_{f=1}^{F} \|\boldsymbol{D}_1^{(f)} - \boldsymbol{D}_2^{(f)}\|. \tag{13}$$

Using eq.(13), the error of our method is defined as the differences between our and the standard models. Moreover, the simulation using the parameter values of a normal liver is made to compare with the proposed method. The obtained models by this simulation are called the comparison models. In Fig. 3 (a), the errors of our method is less than 0.03[mm] while the differences between the standard and the comparison models (dotted lines) are more than 3[mm]. From these results, our method can find the suitable FE parameters compared with the models obtained by the parameter values of a normal liver.

In the second experiment, the proposed method is applied to another 16 test deformation patterns with noise. This is because the measured deformation data include much or less noise due to sensor noise, calibration errors of our system components, and so on. Table. 1 (b) shows the three results of the estimated parameters of the noise patterns by the proposed method. As shown in Fig. 3 (b), the errors of our method (the solid lines) are less than 10[mm]. Also the comparison models are generated by using the parameter values of a normal liver. Comparing with the errors of the comparison models (the dotted lines in Fig. 3), our method is robust against the deformation data with noise.



(a) (b)

Fig. 3. Errors of our method using test deformation patterns with (a)no noise and (b)noise

5 Conclusion

We proposed a method for estimating patient-specific FE parameters by observing the deformation of soft tissues. The proposed method constructs the potential space by sample deformation patterns of a normal liver. Given the deformation of a liver with unknown parameters, the mapped deformation in the potential space is described with a multi-dimensional polynomial curve function. Using the coefficients of the function, the suitable FE parameters for patient's tissue are estimated. From the experimental results, the proposed method can robustly find the reliable FE parameters of the deformation of the tissue with unknown material properties. Now our method have been extended to estimate other FE parameters and boundary conditions.

Acknowledgment. This research is partially supported by Grant-in-Aid for Scientific Research on Innovative Areas, hComputational Anatomy for Computer-aided Diagnosis and Therapyh, MEXT, Japan.

References

1. Fung, Y.C.: Biomechanics: Mechanical Properties of Living Tissues, 2nd edn. Springer (1993)
2. Hong, J., Matsumoto, N., Ouchida, R., Komune, S., Hashizume, M.: Medical navigation system for otologic surgery based on hybrid registration and virtual intraoperative computed tomography. IEEE Transactions on Biomedical Engineering 56(2), 426–432 (2009)
3. Hoshi, T., Kobayashi, Y., Miyashita, T., Fujie, M.: Quantitative palpation to identify the material parameters of tissues using reactive force measurement and finite element simulation. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2822–2828 (2010)
4. Kauer, M., Vuskovic, V., Dual, J., Székely, G., Bajka, M.: Inverse finite element characterization of soft tissues. Medical Image Analysis 6(3), 275–287 (2002)

5. Konishi, K., Nakamoto, M., Kakeji, Y., Tanoue, K., Kawanaka, H., Yamaguchi, S., Ieiri, S., Sato, Y., Maehara, Y., Tamura, S., Hashizume, M.: A real-time navigation system for laparoscopic surgery based on three-dimensional ultrasound using magneto-optic hybrid tracking configuration. International Journal of Computer Assisted Radiology and Surgery 2(1), 1–10 (2007)
6. Matsumoto, N., Hong, J., Hashizume, M., Komune, S.: A minimally invasive registration method using surface template-assisted marker positioning (stamp) for image-guided otologic surgery. Otolaryngology - Head and Neck Surger 140(1), 96–102 (2009)
7. Moffitt, T.P., Baker, D., Kirkpatrick, S.J., Prahl, S.A.: Mechanical properties of coagulated albumin and failure mechanisms of liver repaired using an argon beam coagulator with albumin. J. Biomedical Materials Research (Applied Biomaterials) 63, 722–728 (2002)
8. Morooka, K., Chen, X., Kurazume, R., Uchida, S., Hara, K., Iwashita, Y., Hashizume, M.: Real-time nonlinear FEM with neural network for simulating soft organ model deformation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 742–749. Springer, Heidelberg (2008)
9. Morooka, K., Taguchi, T., Chen, X., Kurazume, R., Hashizume, M., Hasegawa, T.: An efficient construction of real-time fem-based simulator for soft tissue deformation with large dataset. In: CARS 2012, p. 427 (2012)
10. Mousavi, S.R., Khalaji, I., Naini, A.S., Raahemifar, K., Samani, A.: Statistical finite element method for real-time tissue mechanics analysis. Computer Methods in Biomechanics and Biomedical Engineering 15(6), 595–608 (2012)
11. Riken: Human organs property database for computer simulation (2008), `http://cfd-duo.riken.jp/cbms-mp/index.htm`
12. Volonte, F., Pugin, F., Bucher, P., Sugimoto, M., Ratib, O., Morel, P.: Augmented reality and image overlay navigation with osirix in laparoscopic and robotic surgery: not only a matter of fashion. Journal of Hepato-Biliary-Pancreatic Sciences 18(4), 506–509 (2011)

REEM Service Robot: How May I Help You?

Luca Marchionni, Jordi Pages, Jordi Adell, Jose Rafael Capriles, and Hilario Tomé

Pal Robotics SL, Barcelona, Spain
http://www.pal-robotics.com

Abstract. Service robots are a category of robots designed to help people improving their quality of life. REEM is a humanoid service robot able to provide useful applications in public spaces and to assist people in domestic environments. REEM is the result of several years of research in real scenarios. Its main functionalities are autonomous navigation, person detection and recognition, speech synthesis and recognition, object recognition, object grasping and manipulation. In this paper we present these applications and discuss how they are used to deploy REEM in human environments.

Keywords: service robotics, navigation, person detection, speech recognition.

1 Introduction

To improve and to ensure the person's quality of life constitutes an essential task of our society. Service robots are an extensive category of robots which perform services useful to the well-being of humans, which include all non-industrial applications[1].

Two application domains which can include humanoid robots are domestic robots and public relations robots which can have a big impact in the improvement of people daily tasks and in the quality of life. Service robots could indeed serve as companion, assistant, entertainer, guide, dynamic information point, surveillant, telepresence device, housekeeper, waiter, nurse, ...

To be engaged in social scenarios which are compelling and familiar to humans, the robot has to provide a social communicative functionality that is natural and intuitive. This can be achieved if appearance and functionality fit the robots tasks and robots are as self-explaining as possible [1]. REEM is a humanoid service robot developed by Pal Robotics and designed putting special attention in aesthetics and functionalities to make it suitable for in real human environments. Several experiments in shopping malls, hospitals, museums and trade fairs demonstrated that REEM could provide useful applications assisting and entertaining people in public spaces.

The area of service robotics has been in growing interest during last decades. Research is closely related to general indoor navigation, where one of the first

[1] http://www.ifr.org/service-robots/

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 121–130, 2013.

efforts was the Stanford cart [21],[22]. Successive examples are RHINO [24] and Minerva [25], which guided people at museums and exhibitions. One more recent robot is the retail assistant Robovie-II [27], created by ATR in Japan to provide assistance and guidance to elderly shoppers.

Other research groups designed robots for health-care applications. Fraunhofer IPA developed the Care-O-bot[23] mobile robot assistant, which is able to navigate safely in indoor environments, to entertain humans and to assist them thanks to manipulation capabilities. Wasada University presented Twendy One [28], a human symbiotic robot that provides physical supports to the elderly while securing safe physical human robot interaction. CENTRIA created an interactive service robot called Kaveri[26], to assist elderly people and personnel working in elderly houses or in hospitals.

Social service robots in restaurants are starting to appear. A restaurant opened in 2010 in Thailand with robots as waiter-staff[2]. Customers could place orders using user-friendly touch screens and they are served by robots when meals are ready. South Korean robotics company Future Robot Co. Ltd released FURO[3] series which include different selections of robots for restaurant applications such as greeting robot, entertainment robot, order robot for guidance and for taking order, and finally a payment robot to help customers pay after meals.

2 REEM Service Robot

REEM, the service humanoid robot created by PAL Robotics, fits inside domestic and public relations robots and has been designed to provide useful applications in real-world environments. The main restriction which a service robot must accomplish is human safety. For this reason, REEM has been designed with an extensive network of sensors which allows to detect humans and its environment and to react consequently, giving priority to human safety, afterwards to environment protection and finally task accomplishment. Safety restrictions affected REEM's design to include passive safety by means of soft material covers, real-time actuator control, object and people detection, environment awareness, effective HRI and safe wireless communications.

REEM's been designed with a humanoid shape and a nice appearance. It has a compact wheeled base that allows it to move through irregular terrains, a torso and two arms with hands that it uses to interact with its surroundings. The robot is equipped with 26 actuators and numerous sensors. Laser range finders are installed in its base and torso allowing it to navigate and localize with state of the art SLAM techniques robustly, even in crowded places. Additionally REEM is surrounded by ultrasound and infrared sensors that in conjunction with the lasers allow the robot to navigate safely among people. On the head there is a stereo system that gives the robot the functionality to detect and identify places, objects and people. Thanks to stereo microphones also placed in the head the

[2] http://hajimerobot.com/bangkok/en/main.php
[3] http://www.futurerobot.com/contents_eng/sub42.htm

Fig. 1. REEM service robot

robot can be operated by voice. The microphone array also allows to localize the source of commands helping the robot to filter external disturbances and focus its attention on the users.

3 Autonomous Navigation

REEM service robot performs autonomous mapping, localization, planning and obstacle avoidance in large public spaces as shopping malls or trade fairs, where the environment could be crowded and affected by significant modifications over short periods of time [2]. The following sections will provide an overview of the navigation system components developed on REEM.

Mapping. SLAM has been extensively studied in robotics literature. REEM needs a map of the environment to accomplish higher level tasks. While building a metric map of the environment through laser readings processing, REEM learns a model of Wi-Fi signal distribution and captures camera images with their positions in the map. These additional data will be used during localization. To handle large environments (more than 10000 sqm) REEM adopts a multi-mapping approach [13], that consists in dividing the map into sub-maps and building a graph to represent position constraints between them. To further improve robot capabilities we are looking to include semantic labeling [16] of the environment and long term SLAM [14].

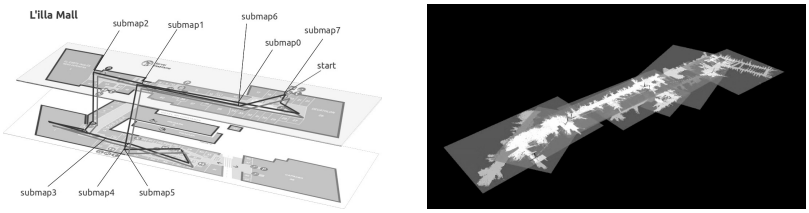


Fig. 2. Multi-mapping at a shopping mall in Barcelona

Localization. Algorithms based on particle filters [15] have been successfully applied in a great variety of robotics applications. Though, sensor readings matching on a previously built map could fail, because the world representation used by the robot won't correspond perfectly to the reality (new obstacles or people) or because of sensor failures (glasses and mirrors). To address this problem REEM monitors the laser-based localization confidence level and performs localization using Wi-Fi signals strength and visual place recognition (databases have been created during map building process). This additional hypothesis help making system more robust and recovering from localization losses.

Path Planning and Obstacle Avoidance. REEM uses a global planner to generate a free path from start to the goal location, based on the previously built map of the environment. Then a local planner modifies the computed trajectory to take into account new obstacles detected by the sensors and to ensure a safe behavior. REEM is able to avoid static and dynamic obstacles updating a costmap with all its sensor readings. To prevent the robot from moving towards dangerous location (stairs), virtual barriers could be specified on the map used by the robot.

To help users in predicting its movements, REEM uses LED indicators and speech synthesis to provide visual and sound feedback of what it is doing. When for example REEM is going to turn left, it moves the head towards that direction and its left ear's LED starts blinking, to show to the users its intentions.

4 Speech Synthesis and Recognition

Speech is the most natural means for human communication and for interact with humans, a robot has to be able to generate and understand speech. REEM has a system that gives it the capability of speech synthesis in several languages. Speaking allows REEM to give help, indicate its intention, warn people around him or give information, in a really natural way.

REEM's microphones are located at each side of the head. We have recently incorporated on REEM an automatic speech recognition (ASR) system that is able to recognize speech commands. Therefore, REEM is not only able to talk to users but also to understand them, for instance understanding command such as *"go to the hall"*.

Current state of the art ASR systems are claimed to reach high accuracies even for large vocabulary continuous speech. Though, accuracy decreases drastically when those ASR systems are applied on a real robot. There are two main reasons for that, both related to the far-talk paradigm: *a*coustic inner-noise and *e*nvironmental interferences. Because the speaker is placed at a certain distance from the microphones which are in the robot, we are in the far-talk framework. This is in contrast to what would be desirable: the user speaking to a microphone in front of the mouth (e.g., a headset). In the far-talk paradigm the speech signal generated by the user attenuates before reaching the microphones. Therefore,

the power of the recorded signal is comparable to the power of robot's inner-noise and external interferences. This issues have to be addressed using specific techniques and represent the current challenges in robust speech recognition for robots.

In order to remove the inner-noise generated by the robot we are currently applying MCRA-based spectral subtraction [5] and for interference removal we are able to remove signals coming from a different direction than the one we are interested in (i.e., user speech). We are basing our algorithms on the work of [6]. Our recent findings indicate that by combining these techniques it is possible to reduce the recognition error by 25% in extreme noisy conditions (i.e., $SNR < 0dB$).

REEM is thus being able to speak to users and understand them, and will be able to do so even in very noisy conditions in the near future.

5 Person Detection and Recognition

Person detection provides the input data to higher level processes like person tracking or behaviors to interact with people. REEM exploits its base laser and its stereo camera in order to detect persons.

Legs Detection. REEM uses the legs detector presented in [7]. This detector consists in an AdaBoost trained to identify what clusters of points are likely to be legs. The detection rate is very high. However, due to the limited information embedded in a laser scan, the number of false detections is pretty high too. The main advantage of leg detection using a laser scanner is that the 3D position is obtained with high accuracy.

Full Body Detection. Our person detection is inspired by [8]. A training dataset has been used consisting in a large number of image windows with non-occluded standing persons and an even larger image windows containing no persons. In order to speed up the detector an AdaBoost cascade based on Haar-like features is firstly used. This cascade is trained to have a high detection rate at a cost of a high number of false alarms. Then, an AdaBoost cascade based on HOG is used to remove false alarms as the discriminative power of this feature is higher. Furthermore, the performance of the latter cascade is improved by using Weighted Linear Fisher Discriminant Analysis (WLFD) in the weak classifiers [9]. The main drawback of person detection in still images is that no 3D position is obtained. Nevertheless, the sight direction in which the person lies is provided. Depth may be roughly estimated by the detection size in the image.

Face Detection. Detecting faces in still images is a well known problem in computer vision [10]. REEM is equipped with a commercial SDK providing both face detection and recognition using monocular images. Face detection is more reliable than full body detection, the latter being more affected by false

alarms. Furthermore, the distance to the person can be inferred by means of stereo triangulation, or by means of monocular vision provided that the eyes are located and the average human baseline is assumed.

Person Recognition. REEM is able to recognize a person by means of matching its face to a database. Therefore, person recognition is restricted to the range of distances in which the face detector is effective. In REEM, provided its cameras and optics, this distance is up to 1.5 meters.

Considering the strong and weak points of each classifier, on REEM, some heuristics are used to provide a more robust detector:

- If there is a face detection and a leg detection, and the directions and distances provided by both detectors are close enough, a person detection with position estimation is provided with the greatest confidence.
- A full body detection whose line of sight is close to the direction of a leg detection and the distance estimated by the vision-based detection size is not far from the one provided by the laser, a high confident detection is provided.
- A full body detection with no matching leg detection provides a medium confident detection.

6 People Tracking

People tracking allows the robot to know the position of a given person, and to update that position while the person or the robot are moving. This skill is used by the robot to follow a person[18] or to reason about surrounding people behavior.

The are several problems to solve: first, make the robot identify each person (i.e. targets) with a unique identifier; second, be able to track this identifier while the person or the robot move; and finally, be robust in front of partial or complete occlusions of the target.

In REEM, people tracking is achieved by a combination of visual and laser features [17]. The people detector described in the previous section uses laser and vision and outputs a list of possible people. Then the people tracker uses a particle filter to track those people, assigning a unique identifier to each of them. The particle filter tracks the position and speed of each person, taking into account current speed and location of the robot for the update step.

To overcome occlusions, two types of sensors are used: laser and vision [19]. This redundancy allows the robot to overcome partial occlusions in one of the other sensors (for example, when the person moves in front of a table, the laser may loose the target, but the vision not). Additionally, the vision system extracts a model that identifies the person uniquely (if completely in view). This model is used by the system to re-target a previously lost target (for example when the target has moved out of the visual scene of the robot, but then reappears).

People tracking has been successfully used on REEM, to make the robot follow a person while the robot is mapping the environment. In such test, the person guides the robot through the environment providing additional information about the place by mean of voice commands. Different techniques for the person following controller have been tested, each one with different advantages or drawbacks [20].

7 Object Recognition and Grasping

REEM has the capability to recognize and estimate the pose of a large variety of textured objects for which a model exists. The model may be a sparse 3D model consisting in a set of 3D points linked to a descriptor or a CAD 3D model where its texture is represented by a sparse set of 3D points and their descriptors. In both cases, SIFT-like descriptors are used to encode the texture appearance of the objects. Regarding how to model the objects, two solutions are considered. Objects that may be modeled as a composition of geometric primitives like cylinders and boxes can be described with a CAD model and its corresponding texture model using the BLORT algorithm [12]. Less structured objects may be also modeled with the 3D sparse model according to the MOPED algorithm [11]. In both cases, key points are first extracted in the image and matches with the texture model are searched. Then, these match candidates are fed into a robust estimation technique like RANSAC in order to solve the Perspective-n-Point (PnP) problem. BLORT goes a step further by validating and refining this pose estimate using a particle filter based on tracking the object inner and outer edges.

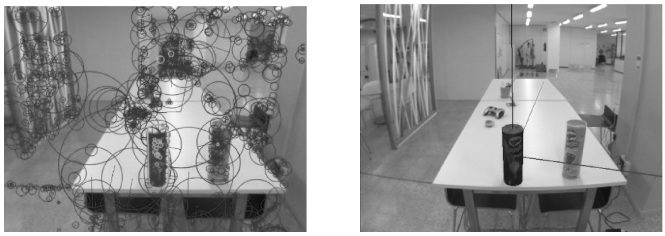


Fig. 3. Left: object detection using texture. Right: object tracking using edges.

REEM's hand has two under-actuated fingers and a thumb. The current control on the fingers allows to estimate the amount of force they are exerting. To grasp an object, after that its pose is known, the robot must generate a motion plan to reach the desired grasp position. The classical approach is using a randomized search algorithm to generate a point-to-point trajectory that brings the arm from its rest position to the desired grasping position. REEM uses the planning framework available on ROS (`http://www.ros.org`).

The generation of this motion plan is computationally expensive and it is very likely that the robot already generated a similar plan in the past. Therefore, we implemented dynamic motion primitives (DMPs) to take advantage from past plans and generate, with negligible computation cost, new motion plans [3]. This is possible as long as the task that they were trained for fits inside the generalization tolerance of the DMP.

In addition, REEM uses visual servoing to overcome calibration errors in the robot's kinematics, and achieve accurate positioning of the hand [4]. The final step is usually with feedback from the hand when the object has actually been grasped, in our case from the current readings of the motors. REEM is capable of grasping common household objects, something that can be very helpful for people that have limited mobility inside their houses.

Fig. 4. Reem acting as guide in two events: very large event IDEX 2011 Abu Dhabi (left), and very crowded event at La Caixa Science Museum Barcelona 2012 (right)

8 Conclusions

REEM robot has been successfully used in several real scenarios where the abilities mentioned were stressed. In all those scenarios, the robot was required to provide information, interact with people and attract them while moving safely around.

From those experiences we conclude that robots like REEM are the future of Service Robots. The reasons are the following:

- Having a human body allows the robot to easily move and interact in a humans world. Furthermore people do prefer a robot that shows a (kind of) human appearance. Every time REEM shows on public places, people starts taking picture and interacting with it. REEM acts as an attractor because of its tall humanoid shape and lively appearance. The human shape is quite of mandatory for appealing people and also for interacting with them in a more familiar way (shaking hands, bowing, etc).
- People want the robots to be autonomous. Instead of telepresence robots or complex teleoperated robots, people prefers robots that are autonomous and decide by themselves. When we use the remote control to control REEM to perform a show in an event, people show deception when they discover

the human behind. Instead, people gets amazed when they observe how the robot moves by itself through the crowd or how it follows a face.
– People prefer robots that look alive. In REEM, face tracking is one of the most successful applications because it shows this liveness by looking at the faces of the people as they move around. Another successful application of REEM is the physical interaction with the robot, like shaking hands. Most people approach the robot either waving their hand to it or requesting it a hand shake, it's like they want to catch the robot attention in order to assess its usefulness.

We think that the future of Service Robots goes through the successful implementation of those three characteristics: humanoid body, autonomy and liveliness. The race to improve those features will mark the agenda of the Service Robots industry for the next years.

References

1. Goetz, J., Kiesler, S., Powers, A.: Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proc. of the 12th IEEE Workshop on Robot and Human Interactive Communication, Roman (2003)
2. Marchionni, L., Tellez, R.: Current challenges in humanoid navigation in dynamic environment. In: Proc. of the IEEE-RAS Proc. Conf. on Humanoid Robots, HUMANOIDS (2011)
3. Lopera, C., Tomé, H., Rodríguez Tsouroukdissian, A., Stulp, F.: Comparing Motion Generation and Motion Recall for Everyday Manipulation Task. In: Proc of the IEEE-RAS Proc. Conf. on Humanoid Robots, HUMANOIDS (2012)
4. Agravante, D.J., Pagès, J., Chaumette, F.: Visual servoing for the REEM humanoid robot's upper body. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA 2013 (2013)
5. Cohen, I., Berdugo, B.: Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. IEEE Signal Processing Letters 9(1), 12–15 (2002)
6. Kim, C., Kumar, K., Stern, R.M.: Binaural Sound Source Separation Motivated by Auditory Processing. In: Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, vol. 4 (2011)
7. Arras, K.O., Mozos, O.M., Burgard, W.: Using boosted features for the detection of people in 2D range data. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA (2007)
8. Paisitkriangkrai, S., Shen, C., Zhang, J.: Real-time pedestrian detection using a boosted multi-layer classifier. In: The 8th Workshop on Visual Surveillance, Marseille, France (2008)
9. Laptev, I.: Improvements of object detection using boosted histograms. In: Proc. of the British Machine Vision Conference, BMVC 2006 (2006)
10. Zhang, C.: Zhang. Z.: A survey of recent advances in face detection. Microsoft Research Technical Report no. 2010-66
11. Collet, A., Martinez, M., Srinivasa, S.: The MOPED framework: object recognition and pose estimation for manipulation. Int. Journal of Robotics Research 30(10), 1284–1306 (2011)

12. Morwald, T., Prankl, J., Richtsfeld, A., Zillich, M., Vincze, M.: BLORT - The Blocks World Robotic Vision Toolbox. In: Best Practice in 3D Perception and Modeling for Mobile Manipulation (2010)
13. Konolige, K., Marder-Eppstein, E., Marthi, B.: Navigation in Hybrid metric-topological maps. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA (2011)
14. Walcott, A.: Long-term Mobile Robot Mapping in Dynamic Environments. MIT PhD Thesis (May 2011)
15. Thrun, S., Fox, D., Burgard, W., Dellaert, F.: Robust Monte Carlo Localization for Mobile Robots. Int. Journal of Artificial Intelligence 128(1-2), 99–141 (2000)
16. Mozos, Ó.M.: Semantic Labeling of Places with Mobile Robots. STAR, vol. 61. Springer, Heidelberg (2010)
17. Bellotto, N., Hu, H.: People tracking with a mobile robot: a comparison of Kalman and particle filters. In: Proc. of the 13th IASTED Int. Conference on Robotics and Applications, RA 2007, pp. 388–393 (2007)
18. Calisi, D., Iocchi, L., Leone, R.: Person Following through Appearance Models and Stereo Vision using a Mobile Robot. In: Proc. of Int. Conf. on Computer Vision Theory and Applications, VISAPP (2007)
19. Choi, W., Pantofaru, C., Savarese, S.: Detecting and Tracking People using an RGB-D Camera via Multiple Detector Fusion. In: Workshop on Challenges and Opportunities in Robot Perception, at the Int. Conf. on Computer Vision, ICCV (2011)
20. Gockley, R., Forlizzi, J., Simmons, R.: Natural person-following behavior for social robots. In: Proc. of the ACM/IEEE Int. Conf. on Human-robot interaction, HRI 2007, pp. 17–24 (2007)
21. Moravec, H.P.: Towards automatic visual obstacle avoidance. In: Proc. of Int. Joint. Conf. on Artificial Intelligence, Cambridge, MA, vol. 584 (1977)
22. Moravec, H.P.: The Stanford cart and the CMU rover. Proc. IEEE 71, 872–884 (1983)
23. Graf, B., Parlitz, C., Hägele, M.: Robotic Home Assistant Care-O-bot® 3 Product Vision and Innovation Platform. In: Jacko, J.A. (ed.) Human-Computer Interaction, HCII 2009, Part II. LNCS, vol. 5611, pp. 312–320. Springer, Heidelberg (2009)
24. Beetz, M., Burgard, W., Cremers, A.B., Fox, D.: Active localization for service robot applications. In: Proc. of the 5th Int. Symp. on Intelligent Robotic Systems, vol. 97 (1997)
25. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Delaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: Minerva: A second generation museum tour-guide robot. In: Proc. of IEEE Int. Conf. on Robotics and Automation, ICRA (1999)
26. Pieska, S., Luimula, M., Jauhiainen, J., Spiz, V.: Social Service Robots in Public and Private Environments. In: Recent Researches in Circuits, Systems, Multimedia and Automatic Control, pp. 190–196 (2012)
27. Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., Kidokoro, N.H.: An Affective Guide Robot in a Shopping Mall. In: Proc. of the 4th ACM/IEEE Int. Conf. on Human Robot Interaction, HRI (2009)
28. Iwata, H., Sugano, S.: Design of human symbiotic robot TWENDY-ONE. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, ICRA 2009, pp. 580–586 (2009)

Training Study Approaches for a SVM-Based BCI: Adaptation to the Model vs Adaptation to the User

Enrique Hortal, Eduardo Iáñez, Andrés Úbeda, José María Azorín, and Eduardo Fernández

Biomedical Neuroengineering Group, Miguel Hernández University of Elche, Av. de la Universidad s/n, 03202 Elche, Spain
{ehortal,eianez,aubeda,jm.azorin,e.fernandez}@umh.es
http://nbio.umh.es

Abstract. Support Vector Machine (SVM) is extensively used in BCI classification. In this paper this classifier is used to differentiate between two mental tasks related to motor imaginary in order to check the possibility of improvement with two alternative adaptation (user's adaptation and model adaptation). Two kind of training have been done by 4 subjects. In the first test (user's adaptation to the model), each subject use a personalized model and 7 sessions are registered to compare the evolution of the results due to the user's training. This initial model is done with a preliminary session which include register of 6 different motor imaginary tasks to select the best combination of them. The second test (model's adaptation to the user) tries to evaluate the benefits of the updating of the model with new registers. The results show that, at least for this kind of imaginary tasks, these methods of adaptation with a SVM-based system have not a meaningful increase of the success rate.

Keywords: BCI, adaptation, SVM, motor imaginary task, training.

1 Introduction

A Brain-Computer Interface (BCI) allows obtaining control commands from the brain information of a person [1]. Traditionally, these kinds of interfaces have been used in order to assist people with disabilities, such as the control of a wheelchair [2], [3] or the management of attendance interfaces [4], [5]. The brain-neural interfaces are also used in motor rehabilitation tasks related to upper limb rehabilitation of patients suffering from CVA (Cerebral Vascular Accident)[6], [7].

Invasive brain interfaces process the neural activity obtained directly from the cerebral cortex [8], [9], but the use of these interfaces in humans involves ethical and medical limitations. Alternatively, the use of non-invasive brain-neural interfaces in humans has been imposed in recent years [10]. In this case, the electroencephalographic (EEG) signals are recorded using electrodes placed on the scalp and they are processed and classified so as to obtain the control commands required.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 131–140, 2013.

The use of non-invasive brain interfaces has been widely studied in relation to the motor activity analysis, Event-Related Potential (ERP) and other evoked potentials such as P300 [11], [12]. In this paper, the goal is the training of a SVM-based system to the detection of two motor imagery tasks utilizing a non-invasive spontaneous interface. A study of two different methods of adaptation is done in this paper. Firstly, the subjects are trained using a unique model using a SVM classifier. The second training is based on the update of the SVM model after every session, using a new register in order to improve the model.

2 Methodology

2.1 Register

The brain interface implemented with the purpose of the classification of two mental tasks consists of a non-invasive spontaneous system. The acquisition of EEG signals is done utilizing 16 (or 8 electrodes in the "User's adaptation to the model" test) electrodes distributed over the scalp. These signals are acquired through the commercial device g.USBAmp of the company g.tec. This device is composed of 16 channels. A frequency sample of 1200 Hz has been used.

The distribution of sensors on the scalp of the subject is as follows (Fig. 1): FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, PO3 and PO4, according to the international 10/20 system [13]. Regarding the placement of the electrodes for conducting the tests, a cap g.GAMMAcap of the company g.tec is used. This cap allows a quick positioning of electrodes without the need of abrasive gel (given the fact that its sensors are active electrodes). The active electrodes are used to diminish the signal-to-noise ratio. Moreover, this system is able to reduce or eliminate motion artefacts and electromagnetic interferences.

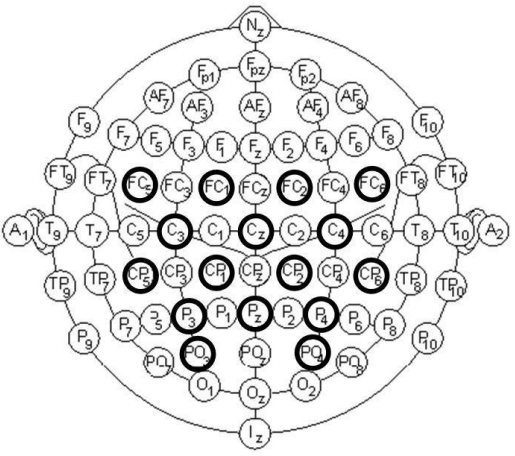


Fig. 1. International 10/20 System

The signals acquired by the amplifier are registered and stored using a computer software developed in MatLab (mathematical software developed by Mathworks). The MatLab API (Application Programming Interface) provided with the device (g.USBAmp MatLab API) is used to manage the software. The subjects are place facing a graphical interface where the different tasks to do are shown. The experimental environment with all the components discussed above is shown in the Fig. 2. The equipment is located in a dedicated room where external stimulus do not disturb the test.

2.2 Processing

Each single EEG signal registered is pre-processed by two different filters. One of them is a notch filter at 50 Hz to eliminate the electrical network noise. In addition, a bandpass filter from 0.1 Hz to 100 Hz is applied to remove the artefacts and the DC component. Both filters are internal hardware filters in the g.USBAmp device. These filters are applied to each electrode and register. As it was discussed above, the frequency of acquisition and stored for the 16 sensors is 1200 Hz.

2.3 Classifier

The last step is to create a classifier that differentiates between two different mental tasks. A classifier based on Support Vector Machine (SVM) has been applied. The SVM classifiers are a very useful technique for data classification [14]. They use a hyperplane or set of hyperplanes in a high or infinite dimensional space to distinguish between different classes.

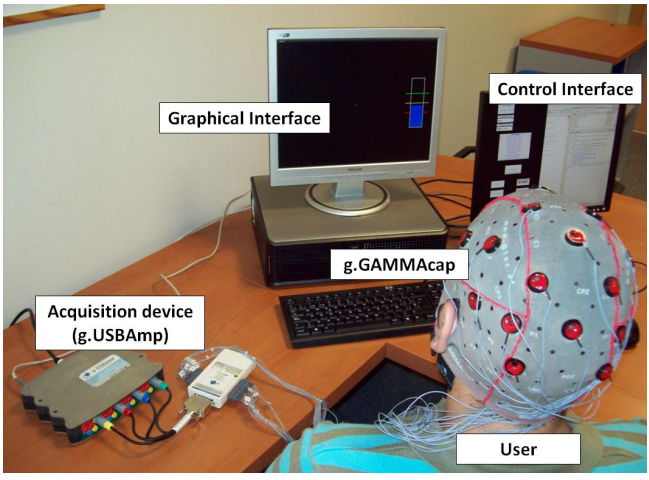


Fig. 2. Experimental environment

The SVM accuracy depends on the kernel selection and the value of their parameters. In BCI classification systems, a Gaussian Kernel or a Radial Basis Function (RBF) [15] are generally used. The two main parameters that configure a RBF kernel are: the parameter of regularization C and the γ parameter that determines the kernel [16], particularly, C = 512 and $\gamma = 0.002$.

The classification procedure consists of using the group of trials obtained from the processing and create two groups, training and test. The aim of SVM is to create a model with the training group, based on the features, that predict the different tasks. Finally, the test group is classified with this model to know the accuracy of the classifier. In this case, the calculated models will be tested with registers in real time.

3 Experimental Procedure

The electroencephalographic (EEG) registers have been done by 4 healthy subjects. These subjects sit in front of a PC screen and they are asked to imagine the performance of motor activity (these activities are described in section 3.1). The paradigm used to the data acquisition protocol [17] is shown in Fig. 3. Firstly, an image of a cross is displayed to inform the user that a cycle is going to start. Afterwards, an image that represents the metal task to imagine is shown during 2 seconds. Finally, a 10 seconds period is established in order to let the subject imagine the mental task. In each register, this process is repeated 8 times for each task, in an aleatory order. There are two different alternative configuration for the "Task procedure" part of the paradigm. The first configuration is used for the offline registers (configuration (a) in Fig. 3). This registers do not have a visual feedback. The second one ((b) in Fig. 3) includes a visual feedback in order to notify the behaviour of the system to the user. The visual feedback is represented for two bars. One of them (the blue bar in Fig. 3 (b)) enlarges when the task detected is correct. The narrowed one (the red bar in Fig. 3 (b)) enlarges if the tasks detected are incorrect.

Previously to the training tests, an initial session is needed. The registers of this session are used in order to select the best combination of tasks (see Sec. 3.1) and to create the model for the user's training (see Sec. 3.2).

After that, two training tests are done. The goal of this test is studying the behaviour of the system with two different trainings. An amount of 14 sessions per user has been registered. These 14 sessions are divided into two groups of 7. Each session is composed by 8 different registers. Each register contains 16 repetitions of the paradigm described in Section 2.1 with two different tasks. The aim of the first 7 sessions is the verification of the improvement in the results when the subject is trained with an invariable model. The last 7 sessions try to verify whether the model is able to adapt to the user when it is updated with new registers.

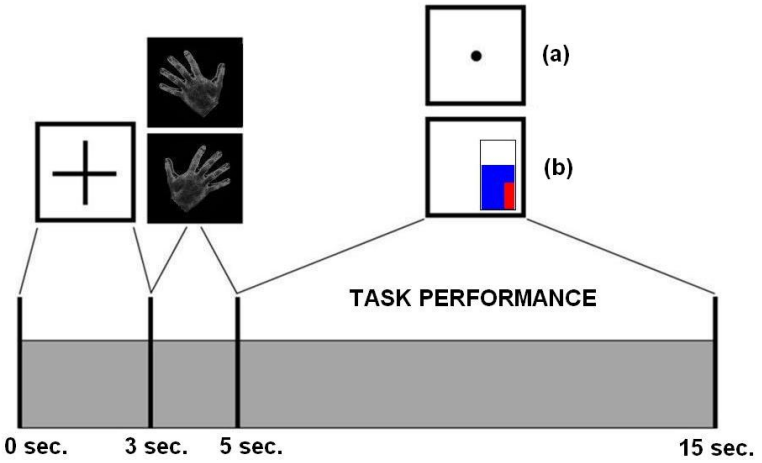


Fig. 3. Paradigm used for the registers. There are two alternative configurations ((a) and (b)). Configuration (a) is used in the offline sessions and configuration (b) is used in the online sessions.

3.1 Tasks Combination Selection

Previously to the process of acquisition of the registers for the training tests, an initial study of the different mental tasks is done (offline sessions). These sessions are done with the configuration (a) in Fig. 3. In this configuration, the user does not have a visual feedback. Taking into account the right and left part of the body, the different combinations of two tasks are studied to find the best of them. The tasks considered in this paper are the following ones:

1. Imagination of the open/close movement of the right hand.
2. Imagination of the open/close movement of the left hand.
3. Imagination of the repetitive low circular movement of the right arm.
4. Imagination of the repetitive low circular movement of the left arm.
5. Imagination of the movement of the right leg.
6. Imagination of the movement of the left leg.

In order to select the best combination of tasks, 4 registers of each pair of tasks are done. Subsequently the success rate from each pair of tasks is obtained. In order to check the results of every combination, a cross validation (using a SVM classifier) is done.

3.2 User's Adaptation to the Model

Firstly, in this paper the results obtained by 4 users in different sessions are compared to check if these users improve the control of the system when they train with an initial model. Before the creation of the model, an offline study is

done to compare the results using each electrode separately. The best 8 electrodes are selected and a model is created with each of them and the decision is taken using the result of each model using the best combination of tasks.

3.3 Model's Adaptation to the User

After the test of the user's adaptation, a new test is done. In this test, a new model is created for each session in order to check whether a modification of the model causes an improvement of the results. Previously, the test to obtain the best electrodes is done again. In all users, an important change in the best electrodes is detected. After the initial training, the results obtained with all the electrodes improve the results with the electrodes separately. For this reason, in this case, the model is done using all the electrodes and a unique model is used to decide between the different tasks. This model is made using the last three sessions of the user and it is updated in each session.

4 Results

In this section, the results of the different tests and the selection of combination of tasks are shown.

4.1 Tasks Combination Selection

At the beginning, a session of 12 registers is done to compare the results of the different combination of tasks used in this paper (see Sec. 3.1). Comparisons between the two groups were made using a cross validation and the results are shown in Tables 1 and 2.

4.2 User's Adaptation to the Model

After the selection of the best combination of task, a model is done following the procedure described in Sec. 3.2. The online results of this test using a 8-fold cross validation is shown in Table 3. In this table are represented the success rate, the error rate, the uncertain rate and the Global Success Rate (GSR in this paper). This parameter represents the success rate without taking into account the uncertain rate.

4.3 Model's Adaptation to the User

In table 4 is shown the results of the second training where the model is update after each session. A 8-fold cross validation is used to obtain the results. In this table the success rate and the error rate are shown.

Table 1. Users 1 and 2: Results of tasks combination selection. Offline success rate %.

Electrode	User 1									User 2								
	1-2	1-4	1-6	2-3	2-5	3-4	3-6	4-5	5-6	1-2	1-4	1-6	2-3	2-5	3-4	3-6	4-5	5-6
FC5	49	54	57	48	48	51	48	48	55	48	44	56	57	49	51	48	53	52
FC1	51	50	52	49	50	51	48	54	52	51	53	74	56	49	49	60	69	49
FC2	48	42	58	52	52	51	62	46	47	60	52	75	60	71	54	64	63	51
FC6	47	51	61	54	54	50	56	45	48	46	53	53	54	46	54	52	51	51
C3	53	45	51	54	55	49	49	57	51	47	59	71	58	73	53	67	58	47
Cz	48	49	63	53	62	51	65	66	45	39	48	48	49	61	51	53	61	53
C4	55	44	52	51	51	49	50	52	46	43	59	66	58	63	55	51	61	46
CP5	47	57	58	51	57	52	54	54	53	56	59	74	57	76	51	66	63	53
CP1	50	45	52	53	56	50	45	49	52	49	59	72	62	76	55	49	66	53
CP2	46	59	49	51	51	53	48	48	51	46	49	71	59	67	50	56	70	50
CP6	51	49	53	52	54	55	51	43	44	50	52	81	58	79	46	78	75	50
P3	47	48	50	48	56	46	40	52	47	48	58	75	61	78	52	58	65	51
Pz	47	43	53	49	48	51	44	56	50	42	52	70	62	77	57	53	69	46
P4	50	47	56	57	53	48	48	47	46	45	58	69	59	77	49	53	65	49
PO3	51	46	50	54	52	44	51	57	55	53	56	64	53	73	53	53	64	52
PO4	51	52	51	48	52	51	54	48	44	49	51	75	56	76	50	54	65	49
Mean	**49**	**49**	**54**	**52**	**53**	**50**	**51**	**51**	**49**	**48**	**54**	**68**	**58**	**68**	**52**	**57**	**64**	**50**

Table 2. Users 3 and 4: Results of tasks combination selection. Offline success rate %.

Electrode	User 3									User 4								
	1-2	1-4	1-6	2-3	2-5	3-4	3-6	4-5	5-6	1-2	1-4	1-6	2-3	2-5	3-4	3-6	4-5	5-6
FC5	53	46	48	51	48	59	58	45	52	56	47	64	56	61	48	48	56	46
FC1	49	46	64	58	50	51	60	66	52	53	46	48	49	52	49	50	46	50
FC2	53	49	51	56	52	53	49	58	50	55	48	50	55	52	52	51	56	55
FC6	48	51	54	50	54	52	51	50	53	51	48	56	52	54	46	60	49	49
C3	52	61	63	52	55	63	57	52	51	56	44	61	56	62	50	59	56	42
Cz	50	48	54	52	62	55	55	50	53	52	53	60	50	53	52	52	53	56
C4	48	47	54	52	51	47	47	55	53	57	50	49	51	51	52	45	48	43
CP5	50	55	58	48	57	46	53	43	49	61	51	51	53	54	56	49	51	50
CP1	57	48	63	51	56	57	56	56	49	48	47	48	48	52	47	54	51	52
CP2	46	46	63	48	51	56	56	60	46	58	57	53	52	57	54	45	50	47
CP6	47	42	47	52	54	54	49	54	52	47	52	53	53	49	49	58	56	53
P3	50	49	56	54	56	56	52	57	52	48	57	51	52	46	59	52	56	49
Pz	57	49	55	58	48	58	58	62	49	54	47	55	54	50	51	57	45	50
P4	57	48	69	52	53	67	66	66	56	50	51	48	49	62	50	53	55	53
PO3	55	48	49	48	52	60	55	54	52	50	48	48	52	45	52	56	53	55
PO4	61	51	52	58	52	56	57	53	54	53	57	50	51	45	50	49	50	50
Mean	**52**	**51**	**56**	**53**	**53**	**56**	**55**	**55**	**51**	**53**	**50**	**53**	**52**	**53**	**51**	**52**	**52**	**50**

Table 3. Results of "User's adaptation to the model" test

Session	Value	User 1 (%)	User 2 (%)	User 3 (%)	User 4 (%)
	Success	37.54	53.54	34.54	38.16
	Error	38.94	34.16	40.01	36.81
	Uncertain	23.52	12.31	25.45	25.03
Session 1	GSR	49.06	61.17	46.25	50.92
	Success	39.43	58.06	36.06	39.83
	Error	38.61	32.77	42.89	34.54
	Uncertain	21.96	9.17	21.05	25.63
Session 2	GSR	50.59	63.75	45.75	53.55
	Success	39.56	59.99	36.19	38.10
	Error	38.03	30.39	39.35	36.97
	Uncertain	22.41	9.62	24.46	24.89
Session 3	GSR	50.50	66.38	47.87	50.76
	Success	38.57	55.39	37.46	37.88
	Error	36.10	32.61	37.83	36.32
	Uncertain	25.33	12.05	24.71	25.80
Session 4	GSR	51.66	62.63	49.70	51.05
	Success	39.31	49.39	36.64	38.46
	Error	36.35	36.10	37.13	35.76
	Uncertain	24.34	14.52	26.23	25.71
Session 5	GSR	51.93	57.75	49.54	51.83
	Success	40.17	58.64	37.21	37.12
	Error	36.27	27.75	38.73	38.07
	Uncertain	23.56	13.61	24.05	24.81
Session 6	GSR	52.57	67.88	48.99	49.38
	Success	38.28	57.12	35.32	39.31
	Error	36.88	27.67	42.27	36.87
	Uncertain	24.83	15.22	22.41	23.82
Session 7	GSR	51.00	67.25	45.55	51.62
	Success	**38.98**	**56.02**	**36.20**	**38.41**
	Error	**37.31**	**31.64**	**39.74**	**36.48**
	Uncertain	**23.71**	**12.36**	**24.05**	**25.10**
Mean	**GSR**	**51.04**	**63.83**	**47.66**	**51.30**

5 Conclusion

The present study was designed to determine the effect of the training of the users and the effect of the update of the model using a SVM classifier. Nevertheless, the results shown in this paper do not improve significantly due to the subject adaptation or the subsequent adaptation of the models.

As future works, further experimental researches are needed to estimate if this study is extended to other imaginary tasks (such as mathematical or arithmetic tasks). Besides, could be very useful repeat this study with different classifiers in order to compare results since a better initial percentages of success. Finally,

Table 4. Results of "Model's adaptation to the user" test

Session	Value	User 1 (%)	User 2 (%)	User 3 (%)	User 4 (%)
	Success	61.60	53.78	55.22	51.56
Session 1	Error	38.40	46.22	44.78	48.44
	Success	54.61	53.99	52.67	50.99
Session 2	Error	45.39	46.01	47.33	49.01
	Success	54.69	51.81	52.88	52.05
Session 3	Error	45.31	48.19	47.12	47.95
	Success	59.09	52.84	55.31	51.97
Session 4	Error	40.91	47.16	44.69	48.03
	Success	62.13	49.88	54.65	52.44
Session 5	Error	37.87	50.12	45.35	47.56
	Success	53.08	51.69	56.50	51.95
Session 6	Error	46.92	48.31	43.50	48.05
	Success	53.95	51.36	57.20	52.93
Session 7	Error	46.05	48.64	42.80	47.07
	Success	**57.02**	**52.19**	**54.92**	**51.98**
Mean	**Error**	**42.98**	**47.81**	**45.08**	**48.02**

this adaptation method will be tested with more healthy users, but also with handicapped people to have a more reliable results.

Acknowledgments. This research has been supported by grant DPI2011-27022-C02-01 of Ministerio de Economía y Competitividad of Spain and by Conselleria d'Educació, Cultura i Esport of Generalitat Valenciana of Spain through grant VALi+d ACIF/2012/135 and grant ACOMP/2013/018.

References

1. Nicolelis, M.A.L.: Actions from Thoughts. Nature 409, 403–407 (2001)
2. Galán, F., Nuttin, M., Lew, E., Ferrez, P.W., Vanacker, G., Philips, J., del Millan, J.R.: A Brain-Actuated Wheelchair: Asyncronous and Non-invasive Brain-Computer Interface for Continuous Control of Robots. Clinical Neurophysiological 119, 2159–2169
3. Iturrare, I., Antelis, J.M., Kubler, A., Minguez, J.: A Non-invasive Brain-Actuated Wheelchair Based on a P300 Neurophysiological Protocol and Automated Navigation. IEEE Transactions on Robotics 25(3), 614–627 (2009)
4. Obermaier, B., Muller, G.R., Pfurtscheller, G.: Virtual Keyboard Controlled by Spontaneous EEG Activity. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 422–426 (2003)
5. Gao, X., Dignfeng, X., Cheng, M., Gao, S.: A BCI-based Environment Controller for the Motion-Disabled. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 137–140 (2003)
6. Wang, C., Phua, K., Ang, K.K., Guan, C., Zhang, H., Lin, R., Chau, K.S.G., Ang, B.T., Kuah, C.W.K.: A feasibility study of non-invasive motor imagery BCI-based robotic rehabilitation for Stroke patients. In: 4th International IEEE/EMBS Conference on Neural Engineering, Ner 2009 (2009)

7. Tam, W.K., Tong, K., Meng, F., Gao, S.: A Minimal Set of Electrodes for Motor Imagery BCI to Control an Assistive Device in Chronic Stroke Subjects: A Multi-Session Study. IEEE Transactions on Neural Systems and Rehabilitation Engineering 19(6), 617–627 (2011)
8. Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M.: Learning to Control a Brain-Machine Interface for Reaching and Grasping by Primates. PLoS Biology 1(2), E42 (2003)
9. Serruya, M.D., Harsopoulos, N.G., Paninski, L., Fellows, M.R., Donoghue, K.: Instant Neural Control of a Movement Signal. Nature 416, 141–142 (2002)
10. Millán, J.R., Ferrez, P.W., Buttfield, A.: Non-Invasive Brain-Machine Interfaces - Final Report. IDIAP Research Institute - ESA (2005)
11. Iáñez, E., Azorín, J.M., Úbeda, A., Ferrández, J.M., Fernández, E.: Mental Task-Based Brain-Robot Interface. Robotics and Autonomous System 58(12), 1238–1245 (2010)
12. Chavarriaga, R., del Millán, J.R.: Learning from EEG Error-related Potentials in Non-Invasive Brain-Computer Interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering 18, 381–388 (2010)
13. American Electroencephalographic Society. American Electroencephalographic Society guidelines for standard electrode position nomenclature. Journal of Clinical Neurophysiology, 8(2), 200–202 (1991)
14. Hsu, C.W., Chang, C.C., Lin, C.J.: A Practical Guide to Support Vector Classification (2003), `http://www.csie.ntu.edu.tw/cjlin/libsvm/` (updated April 2012) (accessed October 11, 2012]
15. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A Review of Classification Algorithms for EEG-based Brain-computer Interfaces. Journal of Neural Engineering 4, R1–R13 (2007)
16. Flórez, F., Azorín, J.M., Iáñez, E., Úbeda, A., Fernández, E.: Development of a Low-cost SVM-based Spontaneous Brain-computer Interface. In: IJCCI (NCTA), pp. 415–421 (2011)
17. Guger, C., Schlögl, A., Neuper, C., Walterspacher, D., Strein, T., Pfurtscheller, G.: Rapid Prototyping of an EEG-Based Brain-Computer Interface (BCI). IEEE Transactions on Rehabilitation Engineering 9(1), 49–58 (2001)

Evaluation of a Telepresence Robot for the Elderly: A Spanish Experience

Javier Gonzalez-Jimenez[1], Cipriano Galindo[1], and Carlos Gutierrez-Castaneda[2]

[1] System Enginnering and Automation Dpt., University of Malaga, Spain
{jgonzalez,cipriano}@ctima.uma.es
http://mapir.isa.uma.es
[2] Distrito Sanitario Costa del Sol, Servicio Andaluz de Salud, Spain
carlosr.gutierrez.sspa@juntadeandalucia.es

Abstract. This paper copes with the evaluation of a robotic telepresence application for the elderly. It presents the experiences on using the Giraff telepresence robot within three real testsites in Spain for more than a year. Results from this evaluation are the basis for future improvements of this technology which very likely will find a field of application in the enhancement of the quality of life of the elderly. This work is framed in the AAL European Joint Programme.

Keywords: Quality of Life Technologies, Robotic Telepresence, Elder Telecare.

1 Introduction

Recent technological advances open uncountable possibilities to improve the quality of life of users, especially those with emotional, physical and/or cognitive impairments. A particular segment of the society that claims a major attention is the elderly, given their dependence and risk of social isolation. In this light, the *Quality of Life Technologies* (QoLTs) concept has emerged to reify the application of different engineering disciplines like computer sciences, robotics, and others to productively improve the daily life of people.

A significant application of the QoLTs is the so-called *robotic telepresence*. Robotic telepresence refers to a combination of technologies that enables a person to be virtually present and to interact in a remote place by means of a robot. Shortly, a person, called *secondary user*, embodies a mobile robot, that is, s/he takes the full control of the robot which physically interacts with the person receiving the service, called *primary user* (see figure 1). The result is that the primary user identifies the robot as the person who is behind (embodied) and establishes a social relation as s/he was actually in the place. A typical scenario where robotic telepresence contributes in the improvement of the quality of life of the elderly is its use by healthcare personnel to regularly visit patients in order to check their general health and mental state.

Robotics in general has striven for the development of assistant robots [13], and concretely in the last years, robotic telepresence, has received great attention

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 141–150, 2013.

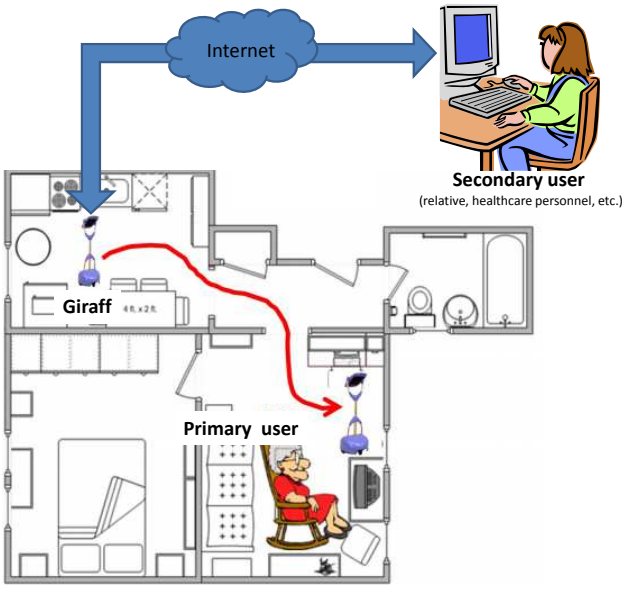


Fig. 1. Actors in a robotic telepresence application. The secondary user, a relative or healthcare personnel, remotely drives the robot deployed in the primary user's apartment and interacts with her through videoconferencing.

from the community, especially applied to the social interaction and assistance of the elderly. Some examples are the Telepresence Robot for Interpersonal Communication (TRIC) [11], Telerobot [10], TeCaRob [12], Mebot [9], etc.

Under this perspective, robotic telepresence is a promising tool that strives for motivating social communication with relatives and friends, helping in mitigating the loneliness of the elderly population, and thus, contributing to improve their quality of life.

A clear example of the concern on developing QoLT is the Ambient Assisted Living (AAL) Joint Programme, an initiative promoted and funded by several AAL Partner States and the European Commission [7]. Under this programme, the work presented here is developed within the project ExCITE (Enabling SoCial Interaction Through Embodiment) [6]. The main objective of this project is the evaluation of the user requirements of social interaction that enables embodiment through robotic telepresence. For that, prototypes of a telepresence robot called *Giraff* (see figure 2a) are deployed at the elders' homes, enabling relatives and healthcare personnel to interact with the primary users.

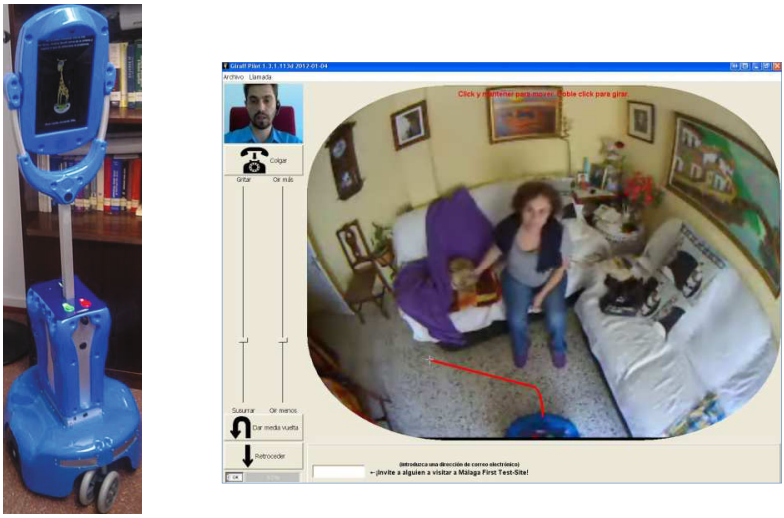
The ExCITE project comes up as an answer to the new generation of senior citizen who are expected to be an active and productive part of the society long after their retirement. The proposed solution considers the robotic telepresence as a manner to improve their quality of life and social interaction. The ExCITE project addresses a Pan-European scale study on the elderly acceptance of this

technology. The study has been carried out in a number of test sites in Sweden, Italy, and Spain in collaboration with healthcare centers, municipalities and research organizations. In this paper we present the experiences and the evaluation of the use of the Giraff robot in a number of test sites in Spain.

The structure of the paper is as follows. Section 2 describes the Giraff telepresence robot. In section 3 the three testsites considered in this evaluation are presented. The general experience and the results of the evaluation is presented in section 4. Finally, conclusions and future work are outlined.

2 The Giraff Telepresence Robot

The Giraff robot, or simply Giraff, is a telepresence robot developed by the Giraff AB company [8]. It consists of a motorized wheeled platform endowed with a videoconferencing set, including camera, microphone, speaker and screen. Giraff permits a virtual visitor to move around, perceive the environment (audio and video), and chat with the primary user. The height of Giraff, the streaming of the visitor camera on the screen, and the possibility of tilting the Giraff's head help in establishing a friendly interaction with the user who can really experience that the visitor is at home.



a) b)

Fig. 2. a) The Giraff telepresence robot. b) The Giraff Pilot application. The secondary user drives Giraff by simply clicking on the image.

From a technical point of view, Giraff relies on a low-cost, commercial computer onboard. The batteries of Giraff last, approximately, two hours and are charged by docking the robot at a station plugged to a normal wall socket of the house.

The Giraff manufacturer (Giraff Technologies AB) provides a software application, called the Giraff *Pilot*, to easily teleoperate the system. Pilot is essentially a graphical interface for driving the robot and controlling the standard videoconference options, i.e., to initiate/hang-up a call, to adjust the speaker and microphone volume, etc. (see figure 2b). At the Giraff side, a program is continually running, accepting calls and providing the necessary functionality for videoconferencing and motion commands. All the actions needed from the elder to handle Giraff can be very easily accomplished with a remote controller. Thus, one of the major advantages of the Giraff telepresence robot is that neither the user nor the visitor need any technological skill to use it, and they both can manage the system (Pilot and Giraff) in an intuitive and natural way.

3 Testsites

Within the context of the ExCITE project, one of the main objectives is to assess the reliability and usefulness of the Giraff telepresence robotic platform as a mean to support elderly and to foster their social interaction.

Rather than testing the system in laboratory setting, the robotic platform is placed in a real context of use. This approach is in line with several research that highlights how systems that work well in the lab are often less successful in real world environments [4]. The evaluation of robots made in a laboratory environment, even though useful, does not favor the emergence of robotic aid suitability to support elders who are able to stay in their own homes. For this reason an essential step is to assess the technology in the specific contexts in which the technology is supposed to be used [5].

Our field evaluation of Giraff is being carried out in three testsites in Spain. Participants were selected by the research team, upon personal interviews with the candidates. The selection criteria took into consideration their interest and motivation, technical and practical requirements of the house, but more importantly it considered their social and healthcare needs. As a result of this process, two women and a man from the province of Malaga (Spain) were selected. They all three are widowers, with ages from 65 to 80, living alone with different grades of dependency. Next, a detailed description of each test site is presented.

3.1 Testsite #1

Testsite #1, localized in the city of Malaga, started in May, 2011. The user is a 65 years old widow lady who lives with one of her sons, though spends a lot of time alone. Though she tries to be occupied in activities organized by the municipality, she misses more personal relations: she only meets some relatives from time to time, and communicates with them by phone on a weekly basis. She is not a technological user, that is, she does not use computers at all, and is not familiarized with videoconferencing either. The main motivation of this person is to be connected to some of their relatives in a warmer and richer manner than a mere phone call.

A specific aim of this testsite is to evaluate the acceptance of a telepresence robot as a companion and communication solution for an active user with an acceptable health and mental state.

The particular layout of this testsite is shown in figure 3. The docking station is placed in the livingroom, where the user spends most of the time, from where visitors can drive Giraff up to the kitchen.

a) b)

Fig. 3. Testsite #1. a) Apartment layout of the rooms covered by the wifi network. The small circle indicates the robot docked at the recharging station. The red trajectory shows the path followed by Giraff in the experience filmed in [2]. b) The elder communicating with a relative.

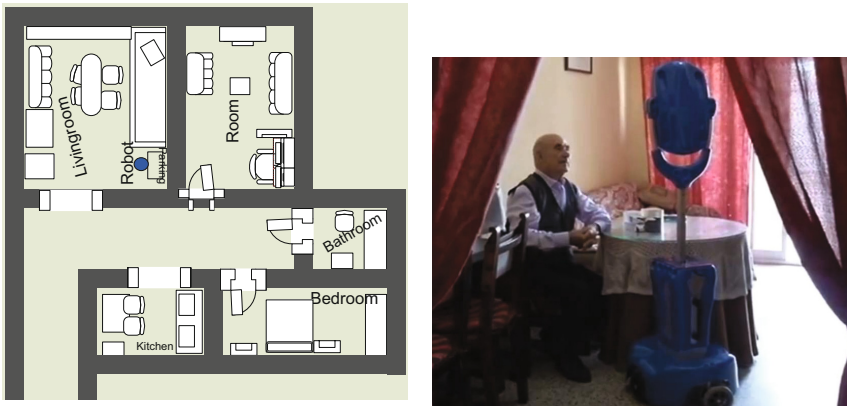
3.2 Testsite #2

This testsite started on November, 2011 in Estepona (Malaga). The primary user is a 80 years old widower who lives alone at home (see figure 4). He is self-sufficient but needs on-site attention by professional health providers, in order to check the evolution of their mental abilities, reviewing the medication, monitoring the blood pressure, temperature, blood sugar level, etc.

In this case, secondary users are, apart from relatives and friends, the professional team of a health center in Estepona, within the Andalusian Public Health Service ("Distrito Sanitario Costa del Sol - Servicio Andaluz de Salud").

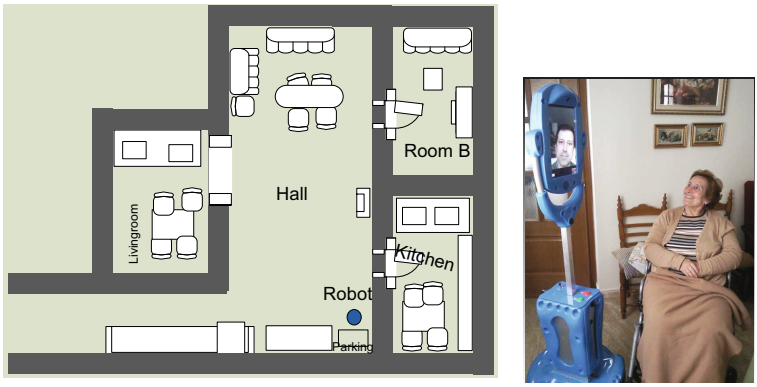
3.3 Testsite #3

This testsite started on January, 2012 in Coín (Malaga) with a 77 years old widow as primary user (see figure 5). Since she is confined in a wheelchair, she is not self-sufficient and needs a caregiver at home 24 hours a day, who does the daily chores and assists her. The main motivation of this person to experience the Giraff telepresence robot is to have a closer contact with her relatives through the Giraff, especially with her sons and grandchildren who live in other cities. The user has not notions of computers and technology whatsoever.



a) b)

Fig. 4. Testsite #2. a) Apartment layout. b) A doctor visits the elder and chats about his medication.



a) b)

Fig. 5. Testsite #3. a) Apartment layout. b) The elder receives the visit of a relative.

4 Evaluation and Experiences

Setting up a testsite involves configuring, at the elderly's home, a pack composed by the Giraff robot, the recharging station, and a remote controller. An obvious, mandatory requirement is the existence of a wifi network with a minimum bandwidth of 6Mb of download speed and 1Mb for uploading. From our experience, we have found that this might be a limitation, at least in Spain where many elder people have not wifi connection at home. In the case of secondary users they only have to install the Pilot software and run it in a computer with internet connection. Given that privacy is a paramount issue in this type of applications,

the secondary user has to be approved to get access to a particular robot as well as to sign in the system with a given credential at every connection.

In order to collect useful feedback of the Giraff telepresence experience from the considered testsites described above we have regularly conducted personal interviews and questionnaires with both, primary and secondary users. They were inquired about suggestions for enhancements in both technical and human-machine interaction features.

4.1 Primary Users

Primary users were asked about their feelings on having a robot at home as a communication mean, as well as about particular problems or inconveniences they have experienced. Results from our evaluations yield a positive view in general. They expressed the technology is easy and valuable, opening for them a new communication channel more interactive and appealing than traditional phone calls. Regarding the acceptance of a robot at home, they consider it as a new appliance that provides a service; the telepresence awareness is enough for them to really "feel" the robot as the person the robot is embodying, especially during long calls.

After our evaluation, the main drawbacks and concerns that primary users find in this technology are:

- The risk of losing the real contact with their relatives could delegate in telepresence instead of physically visiting them.
- Concerns on the electrical consumption of Giraff. The users were worried about this issue, although the normal power consumption of Giraff is reduced, around 100 W. when charging.
- The noise that Giraff robot produces. Internal fans of the robot produces the normal noise of a computer which can be annoying for some people, especially during the night.
- The size of Giraff and the need of placing the docking station in a wide and clear area. This is an inconvenience in small and cluttered houses.
- Impossibility of knowing the caller identity before answering the call.

4.2 Secondary Users

From the secondary-users' point of view, the experience is gratifying in the sense that relatives who live far away from primary users can have the feeling of being in some way closer to them. That could relieve the guilty feeling of not putting enough attention to the elder person, who is living alone. That is particularly important in Spanish and other Mediterranean cultures/societies, where the care of elderly has traditionally been assumed by the family. On the other hand, some of the secondary-users remark a guilty feeling because of the (partial) substitution of real visits for telepresence communication. From the technical side, a frequent complain is related to the quality of the connection which sometimes hinders a fluid communication.

Table 1. Secondary users evaluation. Parking the robot after a visit has been reported as an unpleasant and sometimes annoying task.

Issue	Evaluation
First general impression	4.25
General driving experience	4
General appearance of the interface	4.25
Learning curve	4.75
Moving in straight line	4
Turning	3.75
Parking into the docking station	3.3

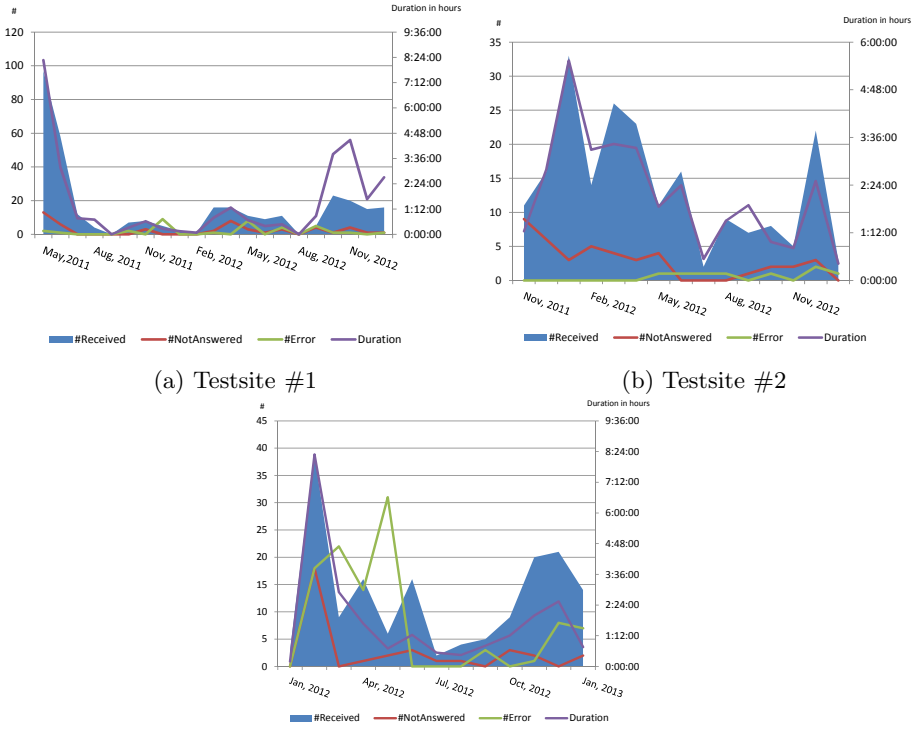
Regarding the ease-of-use of the Giraff robot a number of tests have been conducted with 15 secondary users who have evaluated different aspects of their driving experiences in a scale from 1, (difficult), to 5, (easy). The average scores are shown in the table 1. Note that although the general impressions are positively assessed, most of the secondary users found problems when parking the robot, that is, attaching it to the docking station once the communication has finished. A solution to this issue has been already reported in [1].

4.3 Utilization

During the evaluation presented in this paper, visits' logs have been gathered though a web service provided by Giraff AB. Figure 6 shows the evolution on the utilization of the telepresence robot for each testsite. Notice that at the beginning of the experience the number of received calls and the total duration is higher than in the rest of the period. This is due to the number of testing calls during the setup and also to the initial interest of secondary users to try the technology. As the experience is no longer a something new, the relatives confessed that they sometimes prefer to communicate with the elder via phone rather than turning on the computer and connect to Giraff, and, in contrast, healthcare personnel preferred to maintain telepresence as it is a much better tool for assessment and monitoring health status. Also note the presence of errors in the communication, especially in the case of the testsite #3 (see figure 6c)). These errors, mostly due to problems with the internet connection, have a discouraging effect both for primary and secondary users. Additionally, notice an abrupt decrease in the number of calls and duration during holidays, e.g., around July and August.

5 Conclusions and Discussion

In this paper we have presented an evaluation of the use of a telepresence robot for the elderly. The work presented here, framed in the ExCITE project, illustrates the experiences on three testsites in Spain with different profiles for the primary users, i.e. elders. The testsites, still running, span for almost a year the shortest to more than a year and a half, the longest. The evaluation carried out on them relies on questionnaires and interviews, and has proven the suitability of



(a) Testsite #1

(b) Testsite #2

(c) Testsite #3

Fig. 6. Evolution of the utilization of telepresence for each testsite in terms of number of received, non-answered, and erroneous calls, and duration of effective communication

this technology for improving the quality of life of the elder population, though, of course, there are a number of issues to be addressed in order to make it even more acceptable.

Because of the social relevance of this promising technology, the experiences carried out within this work have caught the attention of the Spanish mass media. Apart from a number of press news, some tv programs have broadcasted this research. Some examples can be found in [2] where the testsite #1 was presented, and in [3] where the testsite #2 was filmed, including, in both, a telepresence connection of a doctor.

Acknowledgements. This work has been supported by the EXCITE project, funded by AAL (Ambient Assisted Living) Program and Instituto de Salud Carlos III.

References

1. Gonzalez-Jimenez, J., Galindo, C., Ruiz-Sarmiento, J.R.: Technical Improvements of the Giraff Telepresence Robot Based on User Evaluation. In: 2012 IEEE RO-MAN: The 21st IEEE Int. Symp. on Robot and Human Interactive Communication (2012)
2. Regional tv show broadcasted (February 2013), http://www.youtube.com/watch?v=D4IgLaqm7ao#t=27m50s
3. Regional tv show broadcasted (May 2012), http://www.youtube.com/watch?v=CXiHhgd_j3g
4. Sabanovic, S., Michalowski, M., Simmons, R.: Robots in the wild: Observing human-robot social interaction outside the lab. In: Proc. of the Int. Workshop on Advanced Motion Control. ACM, Istanbul (2006)
5. Hutchins, E.: Cognition in the Wild. MIT Press (1995)
6. Coradeschi, S., Kristoffersson, A., Loufti, A., Rump, S.V., Cesta, A., Cortellessa, G., Gonzalez-Jimenez, J.: Towards a Methodology for Longitudinal Evaluation of Social Robotic Telepresence for Elderly. In: 1st Workshop on Social Robotic Telepresence, HRI 2011 (2011)
7. AAL Programme webpage http://www.aal-europe.eu/
8. Giraff Technologies AB http://www.giraff.org/
9. Adalgeirsson, S.O., Breazeal, C.: Mebot: a robotic platform for socially embodied presence. In: Proc. of ACM/IEEE Int. Conf. on Human-robot interaction, pp. 15–22 (2010)
10. Michaud, F., Boissy, P., Labonte, D., Corriveau, H., Grant, A., Lauria, M., Cloutier, R., Roux, M., Iannuzzi, D., Royer, M.: A Telementoring Robot for Home Care. In: Mihailidis, A., et al. (eds.) Technology and Aging: Selected Papers from the 2007 Int. Conf. on Technology and Aging, vol. 21. IOS Press (2008)
11. Tsai, Hsu, Y., Ma, A., King, T., Wu, C.: Developing a telepresence robot for interpersonal communication with the elderly in a home environment. Telemedicine and e-Health 13(4), 407–424 (2007)
12. Helal, A., Abdulrazak, B.: TeCaRob: Tele-Care using Telepresence and Robotic Technology for Assisting People with Special Needs. Intl. J. of Human-friendly Welfare Robotic Systems 7(3) (2006)
13. Galindo, C., Cruz-Martin, A., Blanco, J.L., Fernandez-Madrigal, J.A., Gonzalez-Jimenez, J.: A Multi-Agent Control Architecture for a Robotic Wheelchair. Applied Bionics & Biomecanics 3(3) (2006)

Empirical Analysis of the Integration of a BCI and an EOG Interface to Control a Robot Arm

Eduardo Iáñez, Andrés Úbeda, Enrique Hortal, José María Azorín, and Eduardo Fernández

Biomedical Neuroengineering Group, Miguel Hernández University of Elche, Av. de la Universidad S/N, 03202 Elche, Spain
{eianez,aubeda,ehortal,jm.azorin,e.fernandez}@umh.es
http://nbio.umh.es

Abstract. This paper studies the ergonomic issues of a new multimodal application based on the combination of an electrooculography interface and a brain-computer interface to interact with a robot arm. This application allows performing pick and place operations in a specific designed environment. Eight able-bodied volunteers tested the application and were requested to fill a survey about several aspects. To critically discuss and analyze the results, questions about comfort, reliability, usability and usefulness were answered by the participants. The results of the survey show the usefulness of the system and the application, and its relevance when using it with handicapped people with severe neuromuscular disorders. However, the use of multimodality also shows some limitations.

Keywords: ergonomy, multimodal, brain-computer interface, electrooculography, robot arm.

1 Introduction

A human-machine interface (HMI) allows a human to interact with an external device, such as a computer or a robot arm [1, 2]. There are some kinds of HMIs that can be very useful for physically impaired users when used as an assistive technology. For example, electrooculography (EOG) interfaces [3], voice recognition [4] and brain-computer interfaces (BCIs) have been widely used in this sense [2]. Their main goal is to enhance the quality of life of disabled people, increasing their independence and granting greater social inclusion.

The combination of some of these HMIs is called multimodality and can improve the interaction by providing a better control for the user [5, 6]. In this paper, a brain-computer interface and an electrooculography interface have been combined to control a robot arm in a pick and place application. BCIs allow the users to generate control commands with the only help of their thoughts, i.e., without performing any kind of motor movement [7]. There are different kinds of BCIs [8–10]. In this paper, a non-invasive spontaneous BCI has been used. EOG interfaces detect the eye motion by measuring the difference of potential

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 151–160, 2013.

between the cornea and the retina [11]. The EOG interface used in this paper is able to detect the direction of the eye movement (left, right, up and down) and the blink.

There are several works related to multimodality [12, 13]. However, the ergonomic issues involved in the use of these interfaces are not usually studied. The opinion of users is very important to improve the applicability of these interfaces and their ergonomic limitations [14]. In this paper, an empirical evaluation of the multimodal interface developed has been made. A pick and place application that consists of using the multimodal interface to control the actions of a robot arm has been tested. After the tests, a survey with questions about comfort, reliability, usability and usefulness was answered by the eight healthy users that took part in the experiments. The main goal is to obtain a real feedback of the users in order to study the improvements that can be done in the multimodal interface and show if it has benefits regarding control or not. Moreover, the use of brain and ocular interfaces can be of great help for physically disabled users and a substitution of common robot control interfaces such as joysticks or mouses.

The remainder of the paper is organized as follows. In Section 2, the system architecture is presented describing the BCI and the EOG interface and their integration. Section 3 shows the designed application and presents the survey requested to the volunteers. The results and the discussion about the survey are presented in Section 4. Finally, the conclusions are summarized in Section 5.

2 System Architecture

A multimodal interface based on the general diagram shown in Fig. 1 (left) has been designed. This interface uses EOG and EEG signals. To obtain the different signals, the same acquisition device (gUSBamp amplifier from g.tec) has been used. It has 16 channels and several internal filters. The sample frequency used is 1200 Hz. The data are registered each 0.5 seconds for both interfaces and processed independently to obtain two different commands that are combined to get the final output. A bandpass filter between 0.1 and 100 Hz, and a Notch filter of 50 Hz to remove the network noise have been applied. The software used to register the EEG and EOG signals has been programmed in Matlab using the API (Application Programming Interface) provided by the manufacturer (gUSBamp MATLAB API). The processing of the signals and the control architecture has been developed using Matlab. In Fig. 1 (right), a real image of the experimental environment is shown. The user sits in front of the robot arm and manipulates the objects with the help of their thoughts and eye movements.

Brain-Computer Interface. The non-invasive spontaneous BCI designed in [1] has been incorporated into the multimodal interface performing some improvements. Our BCI allows differentiating between three mental tasks related to motor imagery through 7 surface passive electrodes placed on the scalp (before, 16 electrodes were used). As the mental tasks are related to motor imagery, the electrodes have been placed in the following positions of the motor cortex of

the brain: FC1, FC2, C3, Cz, C4, CP1 and CP2, according to the International System 10/20. The users were asked to imagine three different tasks: imagination of a repetitive small circular movement of the left arm; imagination of a repetitive small circular movement of the right arm; and counting operation (backwards from 20) as rest task. A processing has been performed to the EEG signals each 0.5 seconds in windows of 1 second of duration (overlap of 0.5 seconds with the previous sample). A features vector of 56 elements has been obtained combining two algorithms: a FFT between 8 and 16 Hz with a resolution of 2 Hz and a Wavelet Transform selecting three ranges of frequency (0 to 38 Hz, 19 to 28 Hz, and 9 to 19 Hz). The classifier used to differentiate between the mental tasks is based on Linear Discriminant Analysis (LDA). This classifier was developed in [1]. It combines four different models to classify three mental tasks related to motor imagery. A score-based system has been developed to obtain the output of the BCI: right, rest state, left or uncertainty.

EOG Interface. EOG is based on the fact that the eye acts as an electrical dipole between the cornea (positive potential) and the retina (negative potential) [11] (see Fig. 2, left). To obtain the EOG biosignals from the volunteer, four dry flat electrodes have been distributed as it can be seen in Fig. 2, right: two to measure the horizontal channel, two for the vertical channel, and a reference placed on the forehead. The algorithm to detect the eyes movement described in [3, 15] has been improved. The new algorithm is able to detect the gaze direction of the eyes (right, left, up and down) as well as the blink of the volunteers. The volunteers must perform a fast movement of their eyes on the desired direction and then they must return them to the center position. The algorithm works using time windows of 0.25, 0.5 and 1 seconds (previous work worked with windows of 1 second). A window of 0.5 seconds has been used in this work to synchronize the output with the BCI. Another improvement is that the algorithm is even able to detect when an eye movement is performed between two windows of processing. The processing is based on the derivative to detect the abrupt changes of the signal when the volunteer performs a fast movement of their eyes. Now is

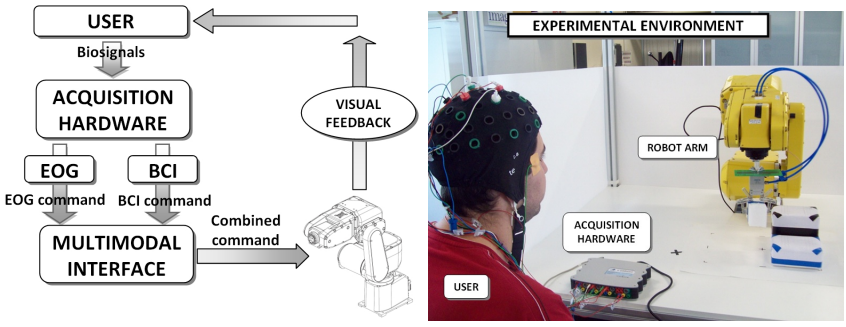


Fig. 1. General diagram of the system (left) and experimental environment (right)

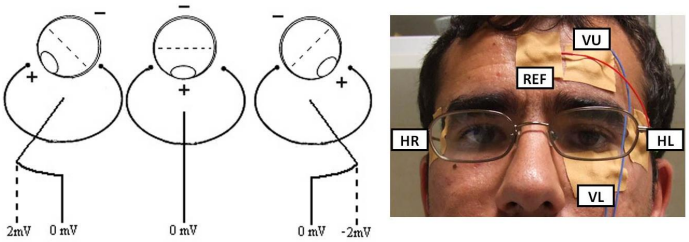


Fig. 2. Ocular dipole (left). EOG electrodes location (right).

able to also detect the blink. To obtain the output, i.e. the direction of the gaze or the blink of the user, each window is analyzed.

Robot Environment. The robot arm used is the Fanuc LR-Mate 200iB model. It has six degrees of freedom and can carry up to 5 kg. A pneumatic gripper has been attached to the end effector of the robot to make possible pick and place operations during the experiments. The robot is placed in a special structure (see right of Fig. 1) with a workspace that makes possible to operate it from a safe distance. It has the proper illumination to perform tasks over a horizontal panel where different objects can be placed to interact with them. The volunteer sits in front of the robot environment. As it can be seen in Fig. 1 (right), two boxes with different height have been placed on the table next to the robot arm. The robot arm can pick and place an object (in this case, another small box) with the pneumatic gripper attached at its end effector.

Integration of Both Interfaces. The BCI and the EOG interface have been combined in order to be able to perform complex tasks, such as pick and place operations in a tridimensional environment using a robot arm. The combination of both interfaces is expected to improve the control of the robot. The robot arm has a gripper attached at its end effector to interact with objects.

Each 0.5 seconds, the EOG and EEG signals are processed to obtain the gaze direction or the blink and the mental task the volunteer is performing. Both commands are combined to interact with the robot arm. On the one hand, the EOG interface is used to control the movement of the robot in the XY plan. On the other hand, the BCI controls the height (Z axis) of the robot. To activate the gripper, the user must perform two consecutive blinks to close or open it depending on the situation. The robot moves 40 mm per second which corresponds to 20 mm per multimodal command sent, as a multimodal command is sent each 0.5 seconds.

This approach is based on a non-simultaneous control of both interfaces which, in terms of usability, makes the operation of the robot easier. When the volunteer performs an EOG command (left, right, up or down), the robot starts moving in the corresponding direction of the XY plane (left, right, forward or backward).

To stop the robot, the user must look into the opposite direction and, then, the robot can be moved again in any other direction. This has been done to make the control more precise. While the robot is stopped, its height can be changed using the BCI. To that end, when the user thinks about the right arm movement, the robot goes up, while it goes down when the mental task performed is the imagination of the left arm movement. A delay of 1.5 seconds has been included when the control changes from the EOG interface to the BCI to avoid undesirable movements. This delay does not affect the decoding of the signal. The brain signals are being processed also during the delay to avoid a bigger delay after the 1.5 seconds. Finally, the gripper is activated using two consecutive blinks. This allows the volunteers to blink normally during the tests to avoid eye fatigue and perform two fast blinks to open or close the gripper depending on the current state.

3 Experimental Tests

The application designed consists of a pick and place operation with objects in an environment using a real robot arm (Section 2). A specific protocol has been followed during the tests to configure both interfaces, perform the training and, finally, do several repetitions of the same experimental sequence. During the execution of the tests, the participants have to complete a pick and place operation that consists of taking an object from its initial position to the top of a box. When the tests are completed, the volunteers are requested to fill a survey about the usefulness, reliability and usability of the system, as well as the effort taken by the users while performing complex tasks.

Eight healthy volunteers (7 male, 1 female) with an average age of 29.9±5.6 (mean±SD) years old took part in the experiments. All volunteers had no medical history of neurological and psychiatric disorders. Five of them needed glasses to obtain a clear vision of the elements of the tests. Six of them were right-handed and the other two were left-handed. After informing the participants of the requirements of the procedure, the volunteers agreed and gave their consents to take part in the tests. All the volunteers are used to work with computers and similar devices (6-10 hours/day). For this reason, they all easily understood the explanation of the tests. One of them was also familiar with BCIs and three more had already participated in previous experiments with similar human-machine interfaces. The users performed the initial training in front of a PC screen at a distance of 1 metre, while the final tests were performed directly in front of the structure of the robot to see clearer the interaction with the objects.

Description of Experimental Tests. Each participant performed the tests in the same day following a particular protocol. First, the necessary equipment to register both EOG and EEG signals is placed. After that, a training with both interfaces individually must be performed. All the process lasts about 80 minutes (including resting periods). After these initial preparations, the volunteers will perform a simple pick and place tasks using the multimodal architecture

designed. In Fig. 1 (right), the appearance of the environment can be seen. The user is sat in front of the robot arm. There are several boxes to simulate an environment with different heights. The small box is going to be moved from an initial position to the top of the other box with the same color. This operation will be performed three times. The robot arm starts in an initial position. The user must use their eye movements to situate the gripper over the small box, and then, he/she must perform the corresponding mental task to approach vertically the robot arm to the box. Finally, he/she must open/close the gripper using their blink. A similar procedure must be followed to place the object over a box. The duration of all three tests, including resting time, lasts about 30 minutes.

Survey Description. After performing the tests, the participants were asked to fill a survey about the use of the multimodal system. First, several questions about personal characteristics, which may be relevant for the study, were made. The second part consisted of the opinion of the volunteers about the tests. The users score each question from 1 (low) to 5 (high). It was divided in five different sections:

- Comfort: questions about the preparation of the tests and equipment. The users were required to answer about comfort issues.
- Reliability: questions about the multimodal interface in terms of accuracy and control. The volunteers were required to answer about the use of the EOG interface and the BCI, and about the combination of both.
- Usability: questions about the time of the tests and the level of stress and tiredness. The volunteers were required to answer about the use of the EOG interface and the BCI, and about the combination of both.
- Usefulness: questions about general satisfaction and helpfulness of the system. The participants were required to answer about the use of the EOG interface and the BCI, and about the combination of both.
- General impressions: finally, some general questions were asked to show the problems and improvements that could be made on the system, particularly in aspects of reliability and future applications.

4 Results

The results obtained from the survey made by all eight users are discussed next. As it has been mentioned before, the survey has been divided in five sections that show a general and particular view of the performance of the tests, comparing the use of both interfaces (ocular and brain) alone or in combination.

Comfort. As expected, the volunteers consider that the equipment used to measure the EOG signals is quite comfortable (3.8 ± 0.9), while the BCI has a normal grade of comfort (2.6 ± 0.9). The placement of the cap and the electrodes is a bit more complex and that was more disturbing for the volunteers. The users

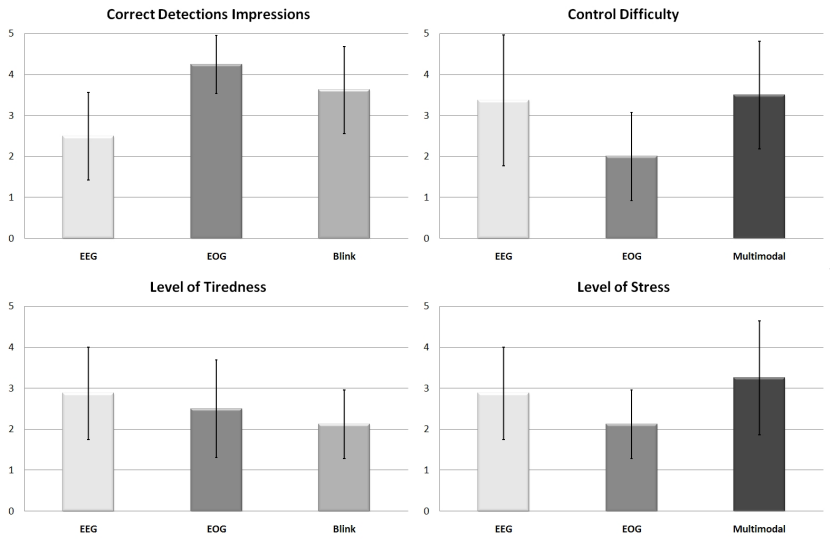


Fig. 3. Survey results of reliability and usability

also indicated that the time of tests was more or less as they expected regarding the application they had to perform, particularly, the use of two different interfaces at the same time. These results are coherent with the experimental design of each session, where the number of tests and time used to train with the interface was limited to avoid tiredness.

Reliability. The accuracy of EOG systems is usually very high. In our case, the success rate obtained is over 95% with a 5% of no-detections (no error) when performing movements. In the case of blinking, the detection is almost a 100%, but a small number of volunteers have similar EOG signals when looking up and blinking, that may lead to no detections of the blink. On the other hand, BCIs are usually less accurate (in this case, 60% average) due to the complexity of the classification of mental tasks. However, the results improved when performing the robot tests, as the volunteers had a more motivating visual feedback.

As it can be seen in Fig. 3 (up-left), the volunteers' impressions about the success rate of the system are, as it was expected, greater for the EOG interface (4.3±0.7) than for the BCI (2.5±1.1). Blinking (3.6±1.1) is also a bit less accurate than the EOG movement. People were also asked about the difficulty on the control of both interfaces and the multimodal combination. Fig. 3 (up-right) shows that the EEG interface (3.4±1.6) is harder to control than the EOG interface (2.0±1.1). According to expectations, this study did not find an improvement using the multimodal combination (3.5±1.3), as the control difficulty in a non-simultaneous control tends to be at least as higher as the one obtained with the most difficult one, in this case, the BCI itself. However, it has been

verified that time can significantly increase without combining both interfaces. The participants also think that changes of the control between interfaces are certainly complex (3.4±1.2). Finally, the volunteers have rated the overall reliability of the system (including the application) with 3.3±0.9, which shows that the whole test is not very difficult to achieve.

Usability. The results of the survey related to the usability of the system show that the tests are not excessively tiring, but performing complex tasks using multimodality is quite stressful. As it can be seen in Fig. 3 (down-left), the level of tiredness during the tests is coherent regarding the design of the test. The BCI needs a constant concentration to generate the commands that allow changing the height of the robot arm, while the EOG movements are performed only in specific moments (when the user changes directions or stops the robot). The blink is only performed twice to pick and place the object activating the gripper.

Tiredness before and after the tests has also been measured, showing that the volunteers started the tests quite relaxed (1.5±0.8) and they were not too tired after about two hours (3.1±0.8). In Fig. 3 (down-right), the stress felt by the participants during the tests is shown. The BCI is more stressful than the EOG interface as it has been already commented. When using the multimodal interface, the stress rises due to the complexity of the operations with the robot. Finally, the impression about the length of the whole session (trainings and tests), which is about 110 minutes, was not considered as a very long period by the volunteers (2.4±1.4).

Usefulness. When the volunteers were asked about their general satisfaction after performing the tests (Fig. 4), most were quite pleased about the EOG interface. However, due to lower accuracy of BCI systems in general, they were not as satisfied with this particular interface, and by extension, with the whole system. Nevertheless, all the volunteers agreed about the usefulness of the system (4.0±0.5).

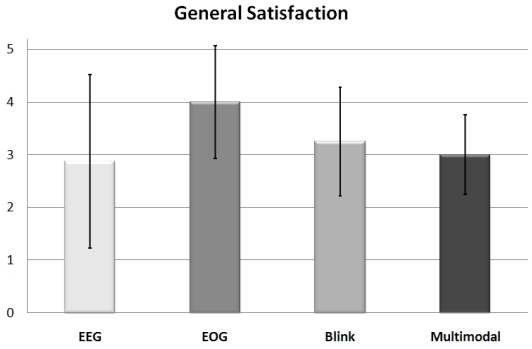


Fig. 4. Survey results of usefulness

General Impression. Finally, several general questions were asked to the participants. They were free to write with their own words their interests and impressions about the multimodal interface proposed.

Several ideas were suggested about improvements of the system. There was a general agreement about the need of enhancing the BCI in terms of accuracy to obtain a better control and avoid stress when thinking. Another volunteer suggested an improvement in the equipment by avoiding the use of conductive gels using other existent acquisition devices that do not need them. This user also proposed to test different mental tasks not necessarily related to motor imagery. About the limitations of the system, there is again a general agreement about the need of avoiding the stress and tiredness provoked by the use of the BCI. Only one user indicated some problems with the EOG blink detection. The multimodality also generates some stress when controlling the robot arm. The speed of the robot movement was also suggested to be increased.

About the general opinion related to future applications of the system, the volunteers remarked that the multimodal interface could be very useful for handicapped people in their daily life to perform simple tasks like using a PC, controlling electrical appliances at home, moving objects or feed themselves.

5 Conclusions

In this paper, an empirical analysis of a multimodal interface based on the combination of a BCI and an EOG interface to control a robot arm has been presented. This multimodal interface allows performing pick and place operations in a tridimensional environment where the height of the robot end effector is controlled through brain activity, its movements in a horizontal plane are handled by ocular commands and the blink closes/opens its gripper. Several tests have been performed by eight healthy volunteers in order to evaluate different aspects of the multimodal interface. To that end, a survey is presented with questions about comfort, reliability, usability, usefulness and general impressions.

As future works, after analyzing the suggestions of the volunteers and the results obtained, an improvement on the integration of both interfaces should be done to solve the limitations remarked by the volunteers. Comfort could be improved by using devices easier to place and that do not require the use of gels. The EOG algorithm will be improved by automatically adjusting the thresholds to prevent errors due to eye fatigue. Finally, this multimodal interface will be tested with more healthy users, but also with handicapped people. This will be important to obtain a real feedback of their special needs. In this sense, it is important to improve the general comfortability of the system and reduce the time taken to perform the tests. It is expected that the use of ocular and brain interfaces, which do not require a physical motor movement to work, could dramatically improve the accessibility of handicapped people to daily tasks such as pick and place objects in a domestic environment.

Acknowledgments. This research has been supported by grant DPI2011-27022-C02-01 of Ministerio de Economía y Competitividad of Spain and by Conselleria d'Educació, Cultura i Esport of Generalitat Valenciana of Spain through grants VALi+d ACIF/2012/135 and ACOMP/2013/018.

References

1. Iáñez, E., Azorín, J.M., Úbeda, A., Ferrández, J.M., Fernández, E.: Mental Tasks-Based Brain-Robot Interface. Robot. Auton. Syst. 58(12), 1238–1245 (2010)
2. de Millán, J.R., Renkensb, F., Mouriño, J., Gerstnerb, W.: Brain-actuated Interaction. Artif. Int. 159, 241–259 (2004)
3. Úbeda, A., Iáñez, E., Azorín, J.M.: Wireless and Portable EOG-Based Interface for Assisting Disabled People. IEEE/ASME T. Mech. 16(5), 870–873 (2011)
4. Monaghan, J., Cheepen, C.: Development and Use of a Speech Recognition System for Physically Handicapped Users. In: Zagler, W.L., Busby, G., Wagner, R.R. (eds.) ICCHP 1994. LNCS, vol. 860, pp. 606–614. Springer, Heidelberg (1994)
5. Leeb, R., Sagha, H., Chavarriaga, R., del Millán, J.R.: A Hybrid BrainComputer Interface Based on the Fusion of Electroencephalographic and Electromyographic Activities. J. Neural Eng. 8(2), 025011 (2011)
6. Zander, T.O., Gaertner, M., Kothe, C., Vilimek, R.: Combining Eye Gaze Input with a Brain-Computer Interface for Touchless Human-Computer Interaction. Int. J. Hum.-Comput. Int. 27(1), 38–51 (2011)
7. Nicolelis, M.A.L.: Actions from Thoughts. Nature 409, 403–407 (2011)
8. Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M., Dimitrov, D.F., Patil, P.G., Henriquez, C.S., Nicolelis, M.A.L.: Learning to Control a Brain-Machine Interface for Reaching and Grasping by Primates. Plos Biol. 1, 193–208 (2003)
9. Obermaier, B., Muller, G.R., Pfurtscheller, G.: Virtual Keyboard Controlled by Spontaneous EEG Activity. IEEE T. Neur. Sys. Reh. 11, 422–426 (2003)
10. Sirvent, J.L., Iáñez, E., Úbeda, A., Azorín, J.M.: Visual Evoked Potential-Based Brain Machine Interface Applications to Assist Disabled People. Expert Syst. Appl. 39(9), 7908–7918 (2012)
11. Barea, R., Boquete, L., Mazo, M., López, E.: Wheelchair Guidance Strategies Using EOG. J. Intell. Robot Syst. 34(3), 279–299 (2002)
12. Lee, E.C., Woo, J.C., Whang, M., Park, K.R.: A Brain-Computer Interface Method Combined with Eye Tracking for 3D Interaction. J. Neurosci. Meth. 190, 289–298 (2010)
13. Iáñez, E., Úbeda, A., Azorín, J.M.: Multimodal Human-machine Interface Based on a Brain-computer Interface and an Electrooculography Interface. IEEE Eng. Med. Biol. Soc., 4572–4575 (2011)
14. Ekandem, J.I., Davis, T.A., Alvárez, I., James, M.T., Gilbert, J.E.: Evaluating the Ergonomics of BCI Devices for Research and Experimentation. Ergonomics 55(5), 592–598 (2012)
15. Iáñez, E., Azorín, J.M., Fernández, E., Sabater, J.M.: Control Strategies for Human-Robot Interaction Using EOG. Int. Symp. Rob., 183–187 (2009)

Interplay between Natural and Artificial Intelligence in Training Autistic Children with Robots

Emilia Barakova[1] and Tino Lourens[2]

[1] Eindhoven University of Technology
P.O. Box 513, Eindhoven, The Netherlands
e.i.barakova@tue.nl
[2] TiViPE
Kanaaldijk ZW 11, Helmond, The Netherlands
tino@tivipe.com

Abstract. The need to understand and model human-like behavior and intelligence has been embraced by a multidisciplinary community for several decades. The success so far has been shown in solutions for a concrete task or a competence, and these solutions are seldom a truly multidisciplinary effort. In this paper we analyze the needs and the opportunities for combining artificial intelligence and bio-inspired computation within an application domain that provides a cluster of solutions instead of searching for a solution to a single task. We analyze applications of training children with autism spectrum disorder (ASD) with a humanoid robot, because it must include multidisciplinary effort and at the same time there is a clear need for better models of human-like behavior which will be tested in real life scenarios through these robots. We designed, implemented, and carried out three applied behavior analysis (ABA) based robot interventions. All interventions aim to promote self initiated social behavior in children with ASD. We found out that the standardization of the robot training scenarios and using unified robot platforms can be an enabler for integrating multiple intelligent and bio-inspired algorithms for creation of tailored, but domain specific robot skills and competencies. This approach might set a new trend to how artificial and bio-inspired robot applications develop. We suggest that social computing techniques are a pragmatic solution to creation of standardized training scenarios and therefore enable the replacement of perceivably intelligent robot behaviors with truly intelligent ones.

Keywords: Artificial intelligence, machine learning, training children with autism with robots, ABA therapy with humanoid robots, bio-inspired computing.

1 Introduction

Developing and validating robot based interventions for children with autism spectrum disorder (ASD) is rapidly growing research field (e.g.,[1,2,3,4,5,6,7,8])

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 161–170, 2013.

which exemplifies the advancement as well as the problems and the needs in combining artificial intelligence, bioinspired methods and human intelligence in real life applications.

Training children with autism with robots is an application domain in which one typically encounters the "dual user" problem: a domain specialist that uses the robot as a tool to augment his/her practice, and the client or patient that is traditionally served by the domain specialist and who is now (partially) served by a robot [5,9]. Therefore, the robot behaviors that are developed for such applications have to serve two aims: enabling and empowering the therapists to train the children, as well as engagement of the robot in natural interaction with the children. Although these aims are very interrelated, they require different robot competencies and features. The empowering the therapists to use robots requires an end-user programming environment, and tools to monitor and easily respond to child behavior through the robot. A natural child to robot interaction requires a robot to have or simulate intelligent behavior and interaction, based on object manipulation and human body language understanding, speech and voice prosody recognition, and robustness to disturbances from human or other environmental agents or circumstances.

Searching for an answer of how behavioral training of autistic children can benefit by the use of robots, we identified that an effective collaboration between therapists, robot specialists, software developers and human robot interaction researchers is a necessity. Each of these communities of researchers or practitioners comes with its own perspective on how robots might be useful and brings its own methodologies and requirements for the use of robots and for the control that humans need to have in this process [9,10]. The of translation and transformation of knowledge between domains could be at the same time beneficial and restricting. For instance, the acceptance that robot can be partially remotely controlled, as in human-robot interaction applications, enables the development of working applications. The actual implementation requires a careful negotiation of the priorities in the development of real-life applications. A typical problems that arise here are to decide the ratio between the robot behaving in a way that it will be perceived as intelligent versus the robot possessing a level of intelligence that drives its behavior. To clarify that point, the perceived intelligence is expressed when the robot behavior is designed to look human-like or clever, but does not require the ability to act according to the interpretation of the sensory information and making behavioral choices based on this and other (for instance memory or learning) information.

Within a multidisciplinary project that includes therapists and child psychologists, robot behavior developers and researchers in human-robot interaction field [5], we analyze the needs, the opportunities, and the balance between including bio-inspired or artificial intelligence methods in application that intend to find a systematic way of including the robots in autism training. We identified 3 groups

of problems where the interplay between natural and artificial intelligence has to be analyzed for challenges and opportunities. These are

1. the empowering and alleviating the work of the therapists, parents and caregivers through a robot based therapy;
2. the creation of tailor-made natural interaction scenarios between a robot and the child;
3. the knowledge translation between the involved domains, i.e. therapists, human-robot interaction specialists and robot behavior developers.

2 Experimental Design as the Basis of the Analysis

In cooperation with experts from Dr. Leo Kannerhuis, a center for autism treatment and research in the Netherlands, three interaction scenarios between a robot and a child with ASD were developed to cover several training sessions for the children. Since the clients of this clinic are high functioning children with ASD, more complex scenarios targeting socialization and communication skills like self-initiation, question asking and problem solving were developed.

During the last four decades techniques and teaching approaches derived from Applied Behavior Analyses (ABA) have been the focus of extensive research for teaching new skills to individuals with ASD [11,12]. Pivotal Response Treatment (PRT) [12] is a promising intervention using principles of ABA. The scenarios that we developed [1], focus for the larger part on pivotal areas [12]. Using standardized ABA training with elements of PRT makes possible that a systematic and gradually more complex training can take place, but the robot scenarios follow several similarly structured training scripts that are part of bigger scenarios and within these scripts variations of intelligent robot behaviors are needed. These include, depending on the scenario, voice prosody and speech recognition, objects, face and gaze following, and providing an appropriate response to the child.

We designed, implemented and carried out three ABA-based robot interventions. The children in these interventions were involved in sessions where they had the opportunity to play game that was mediated by the robot or they interacted with the robot directly. Some of the robot interactions were preprogrammed or autonomous, and at places, when the robot action depended on the answer of the child, the therapist controlled the reaction of the robot by choosing from a number of options [7].

A digital video camera was used to record all sessions. During baseline, intervention with the human trainer and follow-up, several play materials (e.g., LEGO-blocks, toy cars, and plastic toy animals), and pictures were used. The children were familiar with those materials. During intervention sessions with the robot, a NAO robot from Aldebaran robotics was used [1]. The NAO robot is a humanoid robot, however, it has a simple face to prevent confusion or overstimulation in the children. For the voice of the robot a female pre-recorded voice (in one experiment) and text to speech (TTS) computerized voices were used.

3 Robot Intelligence for Empowering the Therapists and Caregivers

In this section we discuss the needs of domain specialists (therapists) in autism training that use the robot as a tool to augment their practice. At this stage of development of robot based therapies the children are always exposed to the robot through or under close supervision of a therapist, parent or a researcher.

From the perspective of the therapists the reliability of the technology is the major issue in these applications. This was the reason to provide an option for remote robot control. A remote controller is also needed to fake intelligent behavior when the creation of such is beyond the achievements of the state of the art research. In addition, this restricts the use of learning algorithms, which have always about 2 percent risk to make a mistake due to the properties of the statistical optimization method [13].

The existing training practices for ASD address specific skills such as: language and communication, problem solving, daily living and appropriate social response. We have chosen ABA, training that utilizes the well-validated psychosocial interventions / training programs of ABA that are clinically employed in ASD and are highly transferable to adaptation for use within robot-based behavioral training scenarios. The ABA scenarios often need to be personalized. In other words, besides the efforts to create useful scenarios, it will be necessary that the therapists change their scenarios "on the fly", i.e. they need to be able to program the robot in an easy and intuitive way.

Initially we believed that creating a sufficient amount of behavioral primitives for the robot will directly enable the therapists to assemble individualized training scenarios for the children. Within the variety of possible interactions, parts of scenarios that included ABA training strategies were repeatedly used. This made us realize, that instead of making the domain specialists use sets of behavioral primitives we can create complete scripts that can be edited for tailored training, while the overall script remains the same. Social computing methods enable the creation of models of complex and interesting social scenarios, possible obstacles that may arise in those scenarios, and potential solutions to those obstacles [14]. The Wikitherapist project [1], aims to make therapists from different clinics develop and share training scenarios that are beneficial for achievement of particular learning goals. Using a community of ABA practitioners we aim to create a set of example scenarios that can be tailored for the particular child by changing some dialog and robot competencies. Such an approach, directed by social computing, will create a unified testbed for share and reuse of robot training scenarios and also enable the faked robot competencies to be resolved with artificial intelligence and bio-inspired methods.

4 The Natural and Artificial Intelligence in Robot-Child Interaction

The challenge in constructing scenario's is seen in the ability of the robot to get engaged in social interaction behaviors such as joint attention, imitation,

turn taking, collaboration. Initiation of social interaction, as a part of ABA training, requires a person to be detected, recognized, or followed. The ability to speak, and recognize even a small set of speech qualities (such as several words and speech prosody), recognize and manipulate some objects, and express social gestures will be very favorable. We will discuss the robot competencies through the needs of the particular ABA therapeutic framework.

4.1 Speech and Voice Prosody Processing

Speech is essential for the interaction between man and a robot and forms a very important aspect in training of autistic children. The NAO robot, used for training, is equipped with Nuance Automatic Speech Recognition (ASR) and Acapela Text To Speech (TTS).

In one of the first experiments conducted in the clinic all spoken sentences were pre-recorded. The motivation to do so was that a dutch TTS system was not yet available on the robot, and spoken text would sound more natural than text converted to computer generated speech.

Recording of natural voices is very time consuming and costly, and small corrections in the text require high implementation effort. When the the dutch TTS became available, it proved to be of acceptable quality. In addition, the TTS synthesized voices provide opportunities for increasing robot autonomy in conversations. The drawback of a TTS system is that it sometimes pronounces words differently, speaks too fast, and sounds somewhat monotonously.

The goal of an ASR system is to accurately and efficiently convert a speech signal into a text message transcription of the spoken words independent of the speaker. Speech recognition is an essential aspect in communication, however automatic speech recognition is a very complex problem, because:

1. despite high recognition rates for a restricted dictionary of over 90 percent (under good circumstances) one has to deal properly with this uncertainty,
2. working with children with delayed speech development and unclear pronunciation increases the chance of fault recognition,
3. the audio recorded by NAO robot contains motor noises, due to positioning of the fans in the vicinity of the microphones, i.e. noise filtering is required.
4. even if the text is properly recognized, it need to be interpreted followed by automatic construction of an appropriate reply, substantial latency might occur resulting in unnatural turn taking in speech interaction.

Voice prosody recognition, which can give a good indication of the persons emotional state is within the reach of the contemporary signal processing and machine learning techniques.

4.2 Visual Data Processing

Face, Gaze, Human Movement and Emotion Recognition. Popular in game engines, like the XBOX in combination with a Kinect (by Microsoft) use

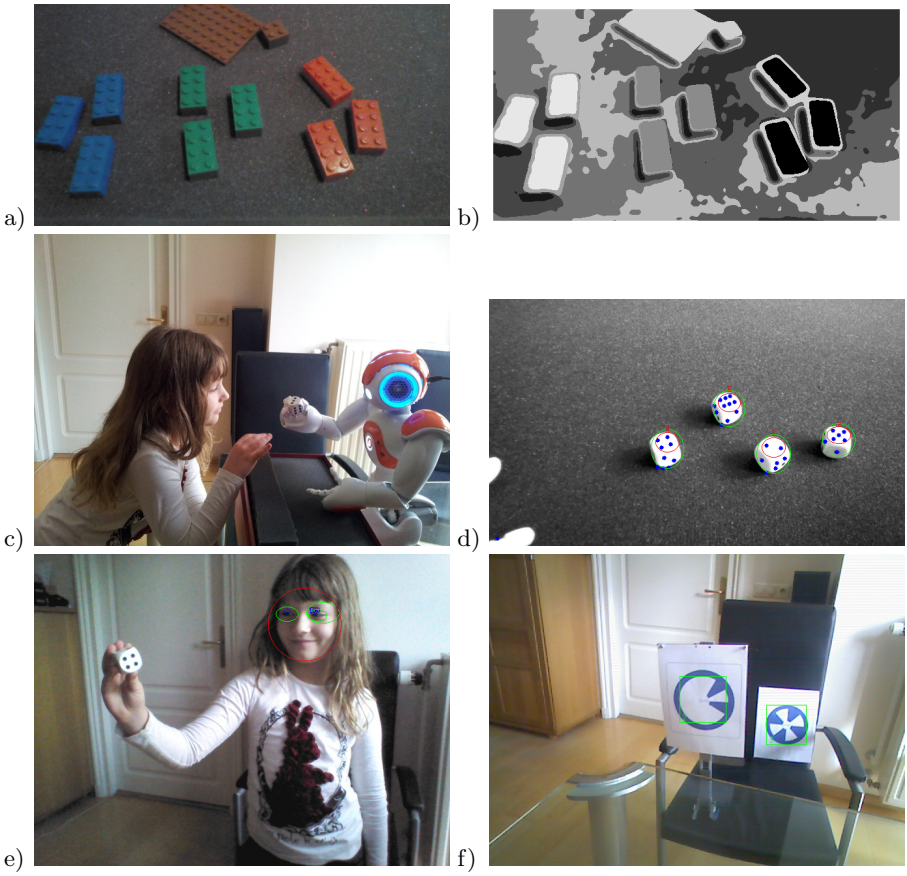


Fig. 1. Vision processing examples, available within TiViPE [15]

skin tones in combination with 3D stereo (infrared laser; time-of-flight) to recognize hands and faces in real time for playing games without using a game controller. These engines have resulted in novel robust recognition methodologies, hand, face, and body posture recognition.

Skin detection in color images by thresholding is an efficient way to locate skin-colored pixels [17,18], which might indicate the existence of human faces and hands. The drawback of this method is that many other objects are have skin-tone colors, such as wood, sand, leather, or fur. Also switching on lights with a warm color (<2700K) makes objects appear skin like, and thus color segregation should always be used in combination with other features. The skin area detection and following has been used for recognition of emotional movements in humans [18].

Generic methodologies to reduce color space dimensionality such as k-means algorithm can facilitate the search for strongly colored objects, as illustrated in Figure 1a,b, represent the input a) and clustered b) image.

In scenario's the robot might be pointing in the direction of the desired color, stating that the child might find the object there. Thus a simple color segregation mechanism might facilitate in scenario development.

Human face recognition has been a topic of extensive research for several decades. Face features are specific by shape and location, because of that databases with thousands of faces available, such that clustering of features can be performed by machine learning algorithms (i. e. on the basis of examples). This yields a robust face recognition systems. The OpenCV software library contains several good algorithms such a Eigenfaces, Fisherfaces, Local binary patterns histograms, or Haar like features can be used. Latter became popular since it can be executed in real time [16]. A Haar-like features, such as edge, line, and center-surround features, considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. For finding the right features a database of images is required, therefore the technique of using Haar-like features is not restricted to faces only, and has been applied to recognize eyes, ears, mouth, and nose, but also lower, upper, and complete body.

Recognition of eyes, and gaze correspondingly, appears to be much more difficult problem, since there are less features in an eye, and different face orientations can hide the eyes for an artificial vision system. For interaction scenarios, however, even one reliable gaze recognition per second or two can contribute to natural interaction Figure 1f.

The quality of face, facial parts, or body recognition strongly depends on the number of (training) examples. Since large databases are available for faces, face recognition works fine using these Haar-like features. For the robot QVGA (320x240) or VGA (640x480) images are used, object recognition works fine withing a 3 meter range, as illustrated in Figure 1f.

Handling Objects for Joint Attention and Turn Taking. Visual data processing is a wide field ranging from early vision features to recognition of complex objects or dynamic features. Applications in object recognition usually target a very specific group of objects, and often require multiple vision features by form, color, motion, and depth. These features can all be extracted from an image, in a generic way. Once these features have been extracted it is still very hard to make up an object, and evaluate when one object is different from another. The way to overcome this problem is to retrieve features from a single object at different views and notify the system that all these views represent the same object, and let the system learn different objects. These object are either learned as a global object or using a leveled approach where component classifiers are used at the first level and the outcome of these classifiers is used to form a so-called single combination classifier [19]. Techniques used are support

vector machines (SVM), cross-correlation in combination with linear classifiers, or feature extraction in combination with a boosting technique [20].

Center-surround, simple, complex, and crossing type of cells, found in the primary visual cortex of mammals [21], resemble similarities with the Haar like features. Bio-inspired operators, [22,23,24] gives in principle similar results as selecting the right Haar feature. The advantage of the bio-inspired operators is that a combination of hierarchies of these features can bring on a long term to the human-like vision system.

Bioinspired operators have been successfully used for handling simple objects like dice for identifying the upper surface of a 3D object and to count the spots (pips) on the top of a dice Figure 1c, d. The target of automatizing a dice game appears to be a relatively simple case of image processing, recognizing pips (Figure 1d) on the top surface of dice is performed using center-surround type of cells. These cells respond strongly to pips,and the recognition rate on these might be well over 99%, but still one needs to consider that falsely recognized pips might lead to heated discussions when playing games with children, therefore the scenario's created always need to have a escape that deals with such a problem.

5 Discussion

We applied a participatory co-creation process to develop real-life robot based behavioral therapies for children with autism. This process combines (a) empowering the therapists with tools to use and reprogram the robot and (b) acquisition of domain specific knowledge from the therapists in order to design robot-child interaction scenarios that accomplish specific learning goals and (c) creation of community of domain specialists and parents that, by the means of social computing methods, will direct the process of scenario creation.

The first two aspects of the process are mutually dependent and comply with the common practice in therapy to create training scenarios, i.e. the therapists define a learning goal that can be achieved through actions eliciting behavior, teaching and practicing skills or providing feedback or encouragement. In attempt to replace part of these training efforts by a robot, we need the robot to acquire gradually more autonomy, which means more currently remotely controlled training actions should be replaced by autonomous actions by the robot. In a search for increasing this autonomy, we have identified repeating structures in the performed ABA based robot scenarios. We found out that it is possible to use ready parts of scenarios that exemplify a training components, and replace the robot interaction behaviors according to the needs and the taste of the individual child. This is to counter the well established approach of using behavioral primitives.

This new approach can only be realized if there is an existence of a common software and hardware platform, that can enable the development of cluster of standardized intelligent behaviors, that can be used interchangeably according to the needs of the particular child. In fact, many individual robot skills and competencies have been developed for a particular robot and language and application,

but still there is the need to reprogram these skills for any new application case. This is one serious setback for the development of general purpose intelligent robot. Since NAO robot has been established as a standardized robot platform, a standardization of the overall platform, including the software and hardware will make possible intelligent solutions to be vastly reused.

In addition, clusters of applications that are useful for the particular application domain are needed. Our findings are based on the work with a single clinic with which we co-developed an ABA-based training tailored for the specific group of children. We have evidence from practice to believe that using the methods from social computing by the community of ABA trained therapists can bring to a creation and sharing of scenarios. While the machine learning methods use examples to learn how to recognize complex objects, the social computing techniques use a community to create knowledge, in our case useful training scenarios, which directs the developments in robot intelligence and interaction.

In summary, we proposed that a common robot and software platform, creation of a cluster of applications, and the use of a community to decide the right directions for the development of the appropriate robot behaviors. We argued that directed, standardized, and truly reusable intelligent behaviors will make possible building on a cluster of applications, and will make possible higher intelligence tasks to be build by combining of simpler robot competencies within standardized scenarios. This could be necessary designers contribution to general robot intelligence that was not achieved for decades.

Acknowledgment. The authors gratefully acknowledge the support of the Innovation-Oriented Research Program Integral Product Creation and Realization (IOP IPCR) of the Dutch Ministry of Economic Affairs, Agriculture and Innovation, The Hague.

References

1. Gillesen, J.C.C., Barakova, E.I., Huskens, B.E.B.M., Feijs, L.M.G.: From training to robot behavior: Towards custom scenarios for robotics in training programs for asd. In: IEEE Int. Conf. on Rehabilitation Robotics (ICORR), pp. 1–7 (2011)
2. Goodrich, M.A., Colton, M., et al.: Incorporating a robot into an autism therapy team. IEEE, Intelligent Systems 27(2), 52–59 (2012)
3. Feil-Seifer, D., Matarić, M.J.: Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In: Khatib, O., Kumar, V., Pappas, G.J. (eds.) Experi. Robotics: The 11th Intern. Sympo., STAR, vol. 54, pp. 201–210. Springer, Heidelberg (2009)
4. Scassellati, B., Admoni, H., Matarić, M.J.: Robots for use in autism research. Annual Review of Biomedical Engineering 14, 275–294 (2012)
5. Barakova, E.I., Gillesen, J.C.C., Huskens, B.E.B.M., Lourens, T.: End-user programming architecture facilitates the uptake of robots in social therapies. Robotics and Autonomous Systems, doi:10.1016/j.robot.2012.08.001
6. Diehl, J.J., Schmitt, L.M., Villano, M., Crowell, C.R.: The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res. Autism Spectr. Disord. 6, 249 (2012)

7. Huskens, B., Verschuur, R., Gillesen, J., Didden, R., Barakova, E.: Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. Journal of Developmental Neurorehabilitation (in press 2013)
8. Wainer, J., Dautenhahn, K., Robins, B., Amirabdollahian, F.: Collaborating with kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. In: 10th IEEE-RAS Humanoids, pp. 631–638 (2010)
9. Goodrich, M.A., Crandall, J.W., Barakova, E.I.: Teleoperation and beyond for assistive humanoid robots. HF Reviews 9 (in press 2013)
10. Kim., E., Paul., R., Shic., F., Scassellati., B.: Bridging the research gap: making hri useful to individuals with autism. J. Hum.-Robot Interact. 1
11. Smith, T., McAdam, D., Napolitano: Autism and applied behavior analysis. In: Sturmey, P., Fitzer (eds.) Autism Spectrum Disorders Applied Behavior Analysis evidence and practice, vol. 13, pp. 1–29. Pro-Ed, Inc. (2007)
12. Koegel, R.L., Koegel, L.K.: Pivotal response treatments for autism: communication, social, and academic development. P.H. Brookes, Baltimore (2006)
13. Barakova, E., Spaanenburg, L.: Windowed active sampling for reliable neural learning. Journal of Systems Architecture 44(8), 635–650 (1998)
14. Boujarwah, F., Abowd, G., Arriaga, R.: Socially computed scripts to support social problem solving skills. In: CHI 2012, pp. 1987–1996. ACM (2012)
15. Lourens, T.: Tivipe –tino's visual programming environment. In: The 28th Annual International Computer Software & Applications Conference, IEEE COMPSAC 2004, pp. 10–15 (2004)
16. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
17. Chai, D., Ngan, K.N.: Face segmentation using skin-color map in videophone applications. IEEE Trans. on Circuits and Systems for Video Techn. 9(4), 551–564 (1999)
18. Lourens, T., van Berkel, R., Barakova, E.I.: Communicating emotions and mental states to robots in a real time parallel framework using laban movement analysis. Robotics and Autonomous Systems 58(12), 1256–1265 (2010)
19. Heisele, B., Serre, T., Pontil, M., Vetter, T., Poggio, T.: Categorization by learning and combining object parts. In: Neural Information Processing Systems (NIPS), Vancouver, pp. 1239–1245 (2001)
20. Heisele, B., Riskov, I., Morgenstern, C.: Components for Object Detection and Identification. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 225–237. Springer, Heidelberg (2006)
21. Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148, 574–591 (1959)
22. Würtz, R.P., Lourens, T.: Corner detection in color images by multiscale combination of end-stopped cortical cells. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 901–906. Springer, Heidelberg (1997)
23. Lourens, T., Barakova, E.I., Okuno, H.G., Tsujino, H.: A computational model of monkey cortical grating cells. Biological Cybernetics 92(1), 61–70 (2005)
24. Lourens, T., Barakova, E.I.: Orientation contrast sensitive cells in primate v1 –a computational model. Natural Computing 6(3), 241–252 (2007)

Use of Multilayer Perceptron vs. Distance Measurement Methods for Classification of Exercises in Telerehabilitation

Oscar Marin-Alonso, Daniel Ruiz-Fernández, Antonio Soriano, and Joaquin D. Garcia-Perez

Dept. of Computer Technology, University of Alicante, Ctra. San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig, Alicante, Spain
{omarin,druiz,asoriano,jdgarcia}@dtic.ua.es

Abstract. Any effort to improve the efficiency of the physical rehabilitation processes is fundamental to ensure the sustainability of healthcare services. This efficiency depends greatly on the patient's adherence to the rehabilitation treatments. Information and communication technologies can help in these issues offering solutions that aim to monitor the patients' rehabilitation exercises performance allowing the existence of domiciliary rehabilitation scenarios. We have developed a solution of this kind, which aims to be as simple and low-cost as possible in the way of how recognizes and evaluates patient's movements. In this work we show a comparison between the use of a multilayer perceptron and a distance between patterns measuring algorithm for patients' motion recognition.

Keywords: telerehabilitation, motion recognition, multilayer perceptron, dynamic time warping.

1 Introduction

Physical rehabilitation processes intend to return patients to a good or healthy condition, state or way of living. Musculoskeletal injuries, certain illnesses, post-surgery situations, or the loss of physical abilities due to aging, are the main reasons for physical rehabilitation processes. These causes can result either in temporary or permanent disabilities. In the first case, rehabilitation plans aim to help the patient regaining maximum self-sufficiency and restore lost skills to live a normal life. If the disabilities are permanent, rehabilitation could last for the whole patient's life with the only goal of maintain his self-sufficiency or even try to slow down the loss of physical skills.

Today, rehabilitation processes have become one of the more demanded healthcare services. World health organization estimates that in 2011 one billion people worldwide were in need of rehabilitation services [1]. In addition, in developed countries this demand is expected to grow due to the increase of life expectancy. The efficiency of these processes is a key point to assure the sustainability of healthcare programmes.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 171–180, 2013.

Talking about physical therapies, efficiency greatly depends on the concept of patients' treatment adherence. Treatment adherence is defined as how well patients follow the physiotherapist's recommendations and the degree of accomplishment of the scheduled clinical activities [2]. The lack of adherence within rehabilitation programmes has been identified as one of the main reasons for treatment failure in physical pathologies [3,4]. Low adherence to treatment can result in longer recovery periods or even in new injuries caused by an erroneous performance of rehabilitation exercises [5], which increase the size of the rehabilitation services' waiting list and affect the patients' quality of life.

Patients' adherence is influenced by factors related to him, the physiotherapist that provides the rehabilitation services and the healthcare organization. An example is the observed difference in treatments' adherence between outpatient and inpatient scenarios [6]. In outpatient cases, adherence tends to decrease throughout the rehabilitation period. Patients' motivation and involvement are critical factors to avoid that decrease and low self-efficacy situations, which are an important barrier to adherence [7]. Studies suggest that a higher effective communication improves motivation and involvement of patients, making the odds of patient' adherence 2.16 higher [8]. In outpatient scenarios the communication between patient and physiotherapist is harder, or even non-existent, which can result in a gap in patients' knowledge about the prescribed exercises, since they perform these exercises without supervision.

Information and communication technologies offer solutions to many of the situations that complicate rehabilitation tasks, in order to make them efficient and assure that healthcare services are affordable to funding institutions.

We have developed a system that aims to help physiotherapists and patients in the management and performance respectively, of outpatient rehabilitation processes. A key issue in this domiciliary scenario is to track and evaluate the patients' performance of the prescribed exercises. This task is based in the comparison of patient's movements with a set pattern.

In this work we show the results of comparing the use of neural networks (specifically, a multilayer perceptron) and an algorithm based on distance measuring between patterns using dynamic time warping (DTW), to evaluate the exercise performance of the patients that use our telerehabilitation system.

In section 2 we show how other researchers previously faced some topics related to our work. In section 3, it is described the telerehabilitation system that we have implemented. Next, in section 4, we describe the experiments performed to evaluate each strategy for motion evaluation: distance measuring between patterns and the use of neural networks. Finaly, in section 5 we show some conclusions of the work carried out and future works related to these conclusions.

2 Related Works

Within the telemedicine research field, telerehabilitation aims to develop and offer a set of techniques and systems to monitor and evaluate the evolution of remotely located patients who follow a rehabilitation plan.

Some works try to offer to patients an innovative way to perform their exercises at home, by means of the use of virtual reality environments. These applications are showed as games, in which the patient has to make a certain set of rehabilitation movements to play. These works are often related to a concrete ailment or set of user, like children rehabilitation or post-stroke situations being difficult to generalize their use [9,10,11]. However, these works barely pay attention to the exercises feedback and the real-time communication of their results to the physiotherapists. The evaluation of the progresses made by the patient takes place regularly at the point of care.

To allow a constant ubiquitous monitoring task, we need to measure, evaluate and communicate the exercises of every rehabilitation session that the patient performs at home. This leads us to the need for motion tracking. In Zhou and Hu survey we can find examples of different tracking systems used for rehabilitation, depending on the sensing technologies and algorithms used [12].

Systems based on inertial sensors, like accelerometers and gyroscopes, are widely used in motion tracking tasks, mainly because of their simplicity and low price. Nevertheless, for rehabilitation purposes it is common to combine them with other sensing devices, in order to overcome inertial sensors' handicaps like drift when estimating positions on 3D space [13,14,15,16,17,18]. These combinations of sensors result in systems that require complex algorithms to process motion data, being also intrusive and annoying for patients. Besides, the cost increases as more sensors and computing capabilities are needed.

Paying attention to cost, simplicity and patients' comfort, we decided to use only one inertial sensing device. We choose Nintendo Wii Remote©, which is a gamepad used with Nintendo Wii© gaming system. This game control includes a three axis accelerometer and uses Bluetooth as wireless connection protocol, with a low cost [19]. Some works show Wii Remote's utility for gesture recognition, given a set of trained gestures and using motion features and machine learning algorithms [20,21,22]. There are few works that have used Wii gaming system with rehabilitation purposes. But they used it just as a tool, not as whole platform to support rehabilitation processes' needs [23,24].

Once motion data is gathered by sensors, should be analysed to recognize the performed movement and evaluate its correctness, duration, intensity etc. Common approaches to do this are the use of Hidden Markov Models [16,20], Extended Kalman and particle filters [13,14], DTW [21,25], and other machine learning methods, including mainly supervised neural networks [16,17]. In all these cases there is a restricted set of movements to be recognized. The work by Hoffman et al. considers recognition up to 25 different gestures being one of the largest studies of this kind [26]. In many cases the purpose of these systems is just to classify a given motion into a set of specific domain of gestures, without evaluating the motion features. In addition, works focused on rehabilitation are usually related to a specific injure in a joint. These are clear drawbacks to generalize the use of these implementations in motion recognition for multiple injuries rehabilitation.

3 eFisiotrack

After having reviewed related works, we found that none of them offered a solution that was real low-cost, handy and intuitive, easy to wear and suitable to rehabilitate different physical conditions in outpatient scenarios. Then, we developed a tool which has all these desirable features.

eFisiotrack is a telerehabilitation software platform to manage and monitor prescribed therapeutic exercises in orthopaedic injuries. It has been conceived to fulfil two primary interlinked requirements: first, to increase patients' adherence to treatment, which helps to satisfy the second, to make rehabilitation processes more efficient and sustainable. Our system tries to offer solutions that improve all the detected issues in telerehabilitation processes discussed before.

Patients using eFisiotrack visit their physiotherapist at the healthcare centre, then the physiotherapist evaluates their physical condition and possible injuries. According to this valoration, the expert schedules a set o exercises for the patient to be done at home.

One of the key features of eFisiotrack is that this platform is designed to deal with a non restricted variety of exercises. That is, eFisiotrack is capable to manage multiple movements in order to rehabilitate whatever joint or part of the patient's body.

Before leaving the physiotherapist's office, the patient peforms each prescribed exercise using the sensing device under the supervision of the physician. His motion is recorded and it will be used as a reference pattern in the motion evaluation process. This is made in order to be as precise as possible evaluating patient's at home performance of the exercises. Moreover, the patterns for each patient and prescribed exercise are expected to vary as they go through their rehabilitation plan, due to their progressive gaining of mobility. This causes a huge increase on the set of motion gestures to be recognized.

The requirements described made us to search for a flexible and easy-to-update motion recognition algorithm. Besides, due to the differences between our system and previous proposals we didn't discard strategies tested in them. From the available options, supervised neural networks and an algorithm based on the use of DTW for measuring the distance between patterns, seemed suitable to solve our needs.

The data gathered by the accelorometer inside the Wii Remote consist on values of linear acceleration along three perpendicular axes (x, y, z). Wii Remote's sensor is able to measure accelerations over a range of ± 3g with 10% sensitivity.

Once the raw acceleration data is transmited to the application, it is filtered using a three point moving average to avoid noisy values on the signal. Then, the filtered acceleration values are the input to both of the motion evaluation strategies tested.

4 Experimentation

To test the suitability of both of the selected strategies, we choose 2 exercises regularly used in shoulder rehabilitation: free active extension and free horizontal

active adduction (Fig. 1). The reason for this choice was two-fold: the prominence of shoulder injuries in physical rehabilitation plans, and the difficulty of clearly recognize these gestures using only acceleration data, due to their similarity.

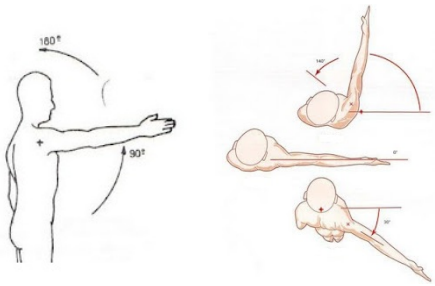


Fig. 1. From left to right, extension and adduction exercises

We built a collection of motion records for each movement. To do this, we selected 5 different users: 3 men and 2 women between 24 and 59 years old. They performed 4 series of 5 iterations for each gesture under a physiotherapist supervision. To obtain two of sets performances, correct and erroneous respectively, they were asked to alternately peform as much precisely as they can and in an more relaxed way each serie of iterations.

4.1 Exercises Classification Using DTW

DTW is an algorithm described in its basics in the work of Sakoe and Chiba [27]. This algorithm was thought to dinamically align time series, in order to estimate the similarity between 2 sequences of values. Sakoe and Chiba conceived it for speech recognition tasks, but since then it has been succesfully used in other fields.

One of the DTW key features for our needs is its ability aligning sequences that differ in duration. This is an important thing for our system, because the length in time of a movement shouldn't be of main importance when estimating the similarity of two sequences. In rehabilitation exercises it is more important that the patient completes the expected trajectory than the time he spends in doing it. Besides, this let us to avoid some pre-processing stages like normalization and quantization, which are compulsory when using other pattern alignment algorithms.

DTW uses a distance function to obtain the best alignment between sequences, which in our case was the well known Euclidean distance. The algorithm estimates the distance between pairs of values of each sequence, allowing the comparison with certain number, warping factor, of neighbour positions in the other sequence. In this multiple comparison, the less distant values are aligned. Finally, the sum of all the distances between aligned values gives a global distance value between the sequences. This value can be interpreted as the similarity degree of those sequences.

We established a threshold empirically estimated. Distance values over this threshold mean that the motion patterns compared are different, so the exercise has been wrongly performed. If the similarity value is lower than the threshold, the exercise is interpreted as correctly executed.

In our experimentation, for each exercise and user we applied our DTW implementation to all the recorded motion sequences, from both of the correct and incorrect sets. We did it 10 times, using in each occasion one correct motion record as a reference sequence. This can bee seen as a leave-one-out cross-validation.

Table 1. Results of gestures recognition of DTW algorithm, in accuracy percentages

Exercise	**Users**										**Total**
	1		*2*		*3*		*4*		*5*		
	✓	*X*	✓	*X*	✓	*X*	✓	*X*	✓	*X*	
Extension	100	94.73	97.89	96.84	100	100	98.94	98.94	98.94	97.89	**97.21**
Adduction	98.94	98.94	97.89	96.84	94.73	94.73	96.84	94.73	97.89	96.84	**96.83**

In Table 1 we show the results obtained in this process, which are expresed in percentages of correct recogniced instances for each user and set of sequences (correct or wrong). The results are highly satisfactory in all cases.

4.2 Exercises Classification Using MLP

Neural networks have been widely applied for motion recognition. In our case we had to implement a classifier to distinguish between 2 classes: motion sequences similar to a correct reference gesture, and motion sequences that are distinct. Taking into account this fact and the availability of 2 sets from correct and incorrect movements, we decided to choose a supervised learning strategie. Concretely we implemented a multi-layer perceptron (MLP) with backpropagation as learning scheme.

The inputs to the designed MLP differ from those used in the DTW case. MLP need a common set of values for every instance to be classified, and for those used in training and validation stages as well. This fact discards the possibility of using the sequences of acceleration values as input to the net. An alternative option is to obtain a set features from the acceleration data. The key issue in doing this is to choose the appropiate set of features, selecting those that combined could offer better classification results.

In our case we selected features related to the field of time series analysis. The set that offered better results contained 11 features:

- Difference between the maximum and the minimum value in the series (peak to peak value).
- Ratio between the largest absolute value of the series and the root mean square (RMS).
- RMS level of the series.
- Root sum of squares level of the series.

- Maximum value of the series.
- Energy of the signal formed by the series.
- Coefficient of correlation between the time series and a reference sequence considered.
- Peak signal to noise ratio.
- Mean square error
- Maximum square error
- Ratio of squared norms.

To those features that need a reference signal (such as the correlation coefficient, the peak signal to noise ratio, the mean square error, the maximum square error and the ratio of squared norms), one of the motion patterns associated to a correct performance was selected.

We estimated the value of the set of features for each axis (x, y, z), having 11 features per axis and making a total of 33 features. We added an extra attribute that has 2 possible values: '0' to identify if the set of features is associated to a right performance, and '1' in case of a wrong performance.

At the end, the dataset used consisted on 15 sets that contain data related to 20 performances (10 correct and 10 performed wrong), one set per user and exercise. For each performance the input sets contain 34 features per axis, making a total of 60 arrays of 34 features for each user and exercise.To train, test and a validate the net we use a 10-fold cross-validation strategie.

Table 2. Results of gestures recognition of MLP implementation, in accuracy percentages

Exercise	**Users**										**Total**
	1		*2*		*3*		*4*		*5*		
	✓	*X*	✓	*X*	✓	*X*	✓	*X*	✓	*X*	
Extension	94.73	84.21	100	89.47	89.47	89.47	94.73	94.73	100	89.47	**92.62**
Adduction	84.21	78.94	89.47	73.68	73.68	73.68	94.73	78.94	89.47	68.42	**80.52**

After an iterative testing process to configure the different MLP's parameters, the best classification results were obtained using a learning rate = 0.3, a momentum = 0.5, training through 150 epochs and having as much hidden layers as input attributes, 34. The results for each exercise, user and kind of performance can be seen in Table 2. The cells contain the percentage of well classified instances in each case.

5 Conclusions

We have developed a project aimed to cover the needs that a physiotherapist and a patient could have through a rehabilitation plan. This is thought to be used in outpatients scenarios, so the monitoring and evaluation of the exercises performance by the patients is crucial.

Paying attention to related works, we noticed that our monitoring, recognizing and motion evaluation needs and requirements differ from previous proposals. We wanted our system to be as simple, fast an easy-to-update as possible. Due to this, we decided to used only one accelerometer as sensing device, basing our motion evaluation strategy only in sequences of accelaration values. Besides this, we didn't want to classify patients performed among a set of gestures. We classify in terms of right performed or wrong performed, comparing acceleration registers from patients performace to a reference register recorded by the patient also, but under physiotherapist supervision.

In this work we make a comparison between 2 strategies for gesture recognition, despite having been tested before. On one hand we use a distance between time series measuring algorithm like DTW. DTW implementation offers an average accuracy rate that ranges from 96% to almost 99% depending on the rehabilitation exercise tested. On the other hand we implemented a supervised neural network, a MLP. MLP implementation average accuracy varys from 80% up to 92% depending on the exercise performed.

Analysing the results from DTW and MLP we notice different issues. DTW shows a smaller variation (1.48%) than MLP (12.1%) in accuracy rates on average, so seems to be consistent independently of the kind of exercise evaluated. This is importan due to the wide variety of exercises the system will include.

Besides, the use of seems more suitable in our system for other reasons. DTW has the drawback of estimate empirically a threshold value to split into 2 classes tue input sequences. MLP does this automatically, but DTW is able to better adapt to differences in lenght between sequences that are compared. As DTW is based in distance mesuring, we can interpret this distance value giving useful information about the patients peformance, e.g., ranging its correctness. In the case of MLP, we can't have access to these data, as we only see a binary output.

Using DTW we work with the original acceleration sequences, which allows us to pay attention to facts like its length. This length can be interpreted as the speed at which the exercise was performed, giving extra information about the patient performance to the experts. For instance, they can deduce patterns of fatigue if the length of the sequences tends to increase as the patient goes through a rehabilitation session.

As future works we consider to incorporate a way to automatically estimate the similarity threshold, based on the performance of each patient, to have a customize version for each one. We also consider to check the relation between acceleration values in each axis to deduce if the motion is taking place in the right plane.

Acknowledgments. This work has been granted by the Ministerio de Economia y Competitividad of the Spanish Government (Ref. TIN2009-10855) and co-financed by FEDER.

References

1. World Health Organization: World report on disability. Technical report, World Health Organization, Geneva (2011)
2. Sabaté, E.: Adherence to long-term therapies: evidence for action. Technical report, World Health Organization, Geneva (2003)
3. Pisters, M.F., Veenhof, C., van Meeteren, N.L.U., Ostelo, R.W., de Bakker, D.H., Schellevis, F.G., Dekker, J.: Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review. Arthritis and Rheumatism 57(7), 1245–1253 (2007)
4. Belza, B., Topolski, T., Kinne, S., Patrick, D.L., Ramsey, S.D.: Does adherence make a difference? Results from a community-based aquatic exercise program. Nursing Research 51(5), 285–291 (2002)
5. Kolt, G.S., Brewer, B.W., Pizzari, T., Schoo, A.M., Garrett, N.: The Sport Injury Rehabilitation Adherence Scale: a reliable scale for use in clinical physiotherapy. Physiotherapy 93(1), 17–22 (2007)
6. Jack, K., McLean, S.M., Moffett, J.K., Gardiner, E.: Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Manual Therapy 15(3), 220–228 (2010)
7. Milne, M.I., Hall, C.R., Forwell, L.: Self-Efficacy, imagery use, and adherence to rehabilitation by injured athletes. Journal os Sport Rehabilitation 14(2), 150–167 (2005)
8. Sniehotta, F.F., Scholz, U., Schwarzer, R.: Bridging the intention-behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 20(2), 143–160 (2005)
9. da Silva Cameirão, M., Bermúdez Badia I, S., Duarte, E., Verschure, P.F.M.J.: Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: a randomized controlled pilot study in the acute phase of stroke using the rehabilitation gaming system. Restorative Neurology and Neuroscience 29(5), 287–298 (2011)
10. Henderson, A., Korner-Bitensky, N., Levin, M.: Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. Topics in Stroke Rehabilitation 14(2), 52–61 (2007)
11. Guberek, R., Schneiberg, S., McKinley, P., Cosentino, F., Levin, M.F., Sveistrup, H.: Virtual reality as adjunctive therapy for upper limb rehabilitation in cerebral palsy. In: 2009 Virtual Rehabilitation International Conference, pp. 219–219. IEEE (June 2009)
12. Zhou, H., Hu, H.: Human motion tracking for rehabilitation-A survey. Biomedical Signal Processing and Control 3(1), 1–18 (2008)
13. Tao, Y., Hu, H., Member, S.: A Novel Sensing and Data Fusion System for 3-D Arm Motion Tracking in Telerehabilitation. IEEE Transactions on Instrumentation and Measurement 57(5), 1029–1040 (2008)
14. Zhang, S., Hu, H., Zhou, H.: An interactive Internet-based system for tracking upper limb motion in home-based rehabilitation. Medical & Biological Engineering & Computing 46(3), 241–249 (2008)
15. Attygalle, S., Duff, M., Rikakis, T.: Low-cost, at-home assessment system with Wii Remote based motion capture. In: 2008 Virtual Rehabilitation, pp. 168–174. IEEE (August 2008)
16. Mannini, A., Sabatini, A.M.: Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. Sensors 10(2), 1154–1175 (2010)

17. Khan, A.M., Lee, Y.K., Lee, S., Kim, T.S.: Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly. Medical & Biological Engineering & Computing 48(12), 1271–1279 (2010)
18. Schönauer, C., Pintaric, T., Kaufmann, H.: Full body interaction for serious games in motor rehabilitation. In: Proceedings of the 2nd Augmented Human International Conference on - AH 2011, pp. 1–8. ACM Press, New York (2011)
19. Lee, J.C.: Hacking the Nintendo Wii Remote. IEEE Pervasive Computing 7(3), 39–45 (2008)
20. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a Wii controller. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction - TEI 2008, p. 11. ACM Press, New York (2008)
21. Leong, T.S., Lai, J., Panza, J., Pong, P., Hong, J.: Wii Want to Write: An Accelerometer Based Gesture Recognition System. In: International Conference on Recent and Emerging Advanced Technologies in Engineering, pp. 4–7 (2009)
22. Kratz, S., Rohs, M., Laboratories, D.T.: A $3 gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors. In: Proceedings of the 15th International Conference on Intelligent User Interfaces, Hong Kong, pp. 341–344 (February 2010)
23. Loureiro, R., Valentine, D., Lamperd, B., Collin, C., Harwin, W.: Gaming and Social Interactions in the Rehabilitation of Brain Injuries: A Pilot Study with the Nintendo Wii Console. In: Langdon, P.M., Clarkson, P.J., Robinson, P. (eds.) Designing Inclusive Interactions, pp. 219–228. Springer, London (2010)
24. Fung, V., Ho, A., Shaffer, J., Chung, E., Gomez, M.: Use of Nintendo Wii Fit™ in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial. Physiotherapy 98(3), 183–188 (2012)
25. Raso, I.: M-Physio: Personalized accelerometer-based physical rehabilitation platform. In: UBICOMM 2010: The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, pp. 416–421 (2010)
26. Hoffman, M., Varcholik, P., LaViola, J.J.: Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices. In: 2010 IEEE Virtual Reality Conference (VR), pp. 59–66. IEEE (March 2010)
27. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 26(1), 43–49 (1978)

Application of Artificial Metaplasticity Neural Networks to Cardiac Arrhythmias Classification

Y. Benchaib[1], Alexis Marcano-Cedeño[2], Santiago Torres-Alegre[2], and Diego Andina[2]

[1] Department of Biomedical Electronics, UABB Tlemcen, 13000, Algeria
benchaib@mail.univ-tlemcen.dz
[2] Group for Automation in Signals and Communications, Technical University of Madrid, Madrid, 28040, Spain
amarcano@gbt.tfo.upm.es, santiago.torres@upm.es, andina@gc.ssr.upm.es

Abstract. Correct diagnosis of cardiac arrhythmias is one of the major problems in medical field. Cardiac arrhythmias can be early detected and diagnosed to prevent the occurrence of heart attack as well as the consequent deaths. An effective method for early detection of these arrhythmias, and thus to procure early treatment, is necessary. In this research we have applied artificial metaplasticity multilayer perceptron (AMMLP) to cardiac arrhythmias classification. The MIT-BIH Arrhythmia Database was used to train and test AMMLPs. The obtained AMMLP classification accuracy of 98.25%, is an excellent result compared to the classical MLP and recent classification techniques applied to the same database.

1 Introduction

Cardiovascular diseases (CVDs) also called heart diseases are a group of disorders of the heart and blood vessels [1]. They are the mayor cause of death worldwide as stated by the World Health Organization [1]. An estimated 17.3 million people died from CVDs in 2008, representing 30% of all global deaths. Out of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke. During 2011 and 2012, 44.5% of deaths in Algeria were due to heart disease, surpassing deceases caused by road accidents. By 2030, almost 25 million people will die from CVDs, mainly from heart disease and stroke, remaining as the single leading cause of death [1]. 7.5 million deaths each year, or 13% of all deaths can be attributed to raised blood pressure. This includes 51% of deaths due to strokes and 45% of deaths due to coronary heart disease [1]. Recent studies show that generally there are significant cardiovascular abnormal symptoms such as palpitations, faints, chest pain, shortness of breath etc., which appear before the sudden occurrence of a heart attack. If these abnormal symptoms could be early detected and diagnosed, time would be saved to prevent the occurrence of heart attack or to provide an efficient treatment in time [2]. Therefore, to reduce the number of disabilities and deceases caused by heart attack, it is necessary to have an effective method for early detection and treatment [2].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 181–190, 2013.

Different artificial neural networks (ANNs) have been suggested for the detection of cardiac arrhythmias. Jadhav proposed a multilayer perceptron (MLP) feedforward neural network model with static backpropagation algorithm to classify arrhythmia cases into normal and abnormal classes. The experimental results gave 86.67% testing classification accuracy [3]. Dayong Gao et al., presented a diagnostic system for cardiac arrhythmias from ECG data, using an ANN classifier based on a Bayesian framework obtaining 90% classification accuracy [4]. Himanshu Gothwal [5] presents a method to analyze electrocardiogram (ECG) signal combining Fast Fourier Transform and neural network trained with Levenberg Marquardt Back-Propagation algorithm, best result obtained was 98.48% for classification accuracy. Sung Nien Yu et al., [6] combines independet componet analysis (ICA) an neural networks obtaining 98.71% classification accuracy.

In this research we apply the artificial metaplasticity multilayer perceptron (AMMLP) algorithm for the classification of different cardiac arrhythmias. The proposed training algorithm is inspired by the biological metaplasticity property of neurons and Shannon's information theory. This algorithm is applicable to ANNs in general, although here it is applied to a MLP. AMMLP algorithm were tested using the well-known MIT-BIH dataset. For assessing this algorithm's accuracy of classification, we used the most common performance measures: specificity, sensitivity and accuracy. The results obtained were validated using the 10-fold cross-validation method.

The remainder of this paper is organized as follows. Section 2 presents a detailed description of the database and the algorithms. The experimental results obtained are present in Section 3. A brief discussion of these results is showed in Section 4. Finally section 5 summarizes the main conclusions.

2 Materials and Methods

2.1 ECG Holter

ECGs are very widely used as inexpensive and noninvasive means of observing the physiology of the heart. By examining the sequence of events on the ECG cardiologists are able to diagnose cardiac arrhythmias. One of cardiac surveillance techniques based on ECG monitoring is the HOLTER monitoring. In 1961, Holter [7] introduced techniques for continuous recording of the ECG in ambulatory subjects over periods of many hours; the long-term ECG (Holter recording), typically with a duration of 24 hours, has since become the standard technique for observing transient aspects of cardiac electrical activity. ECGs are used to diagnose heart disease, to identify irregular cardiac rhythms (arrhythmias), to evaluate the effects of drugs, and to monitor surgical procedures. The magnitude, conduction, and duration of these potentials are detected by placing electrodes on the patient's skin. From the ECG tracing, the following information can be determined: Heart rate; Heart rhythm; Conduction abnormalities (abnormalities in the way the electrical impulse spreads across the heart); Coronary artery disease; Heart muscle abnormality, etc. [8].

2.2 MIT-BIH Dataset

The MIT-BIH Arrhythmia Database was the first generally available set of standard test material for evaluation of arrhythmia detectors, and it has been used for that purpose as well as for basic research into cardiac dynamics at about 500 sites worldwide since 1980 [9]. The group of research in MIT began collecting, digitizing, and annotating long-term ECG recordings obtained by the Arrhythmia Laboratory of Boston's Beth Israel Hospital (BIH).

Database contains 48 half-hour excerpts of two-channel, 24-hour, ECG recordings obtained from 47 subjects (records 201 and 202 are from the same subject) studied by the BIH Arrhythmia Laboratory. These 48 half-hour excerpts were split in two groups: 23 (the "100 series") were chosen at random from a collection of over 4000 Holter tapes, and the other 25 (the "200 series") were selected to include examples of uncommon but clinically important arrhythmias that would not be well represented in a small random sample. The subjects included 25 men aged 32 to 89 years and 22 women aged 23 to 89 years; approximately 60% of the subjects were inpatients.

In most records, one channel is a modified limb lead II (MLII), obtained by placing the electrodes on the chest as is standard practice for ambulatory ECG recording, and the other channel is usually V1 (sometimes V2, V4, or V5, depending on the subject) [10]. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database [9].

2.3 Data Preparation

From 109,871 annotated heartbeats (ECG beats examined by specialists in MIT-BIH), 1000 were selected for this study, which contain 4 different waveforms related to cardiac arrhythmias target. The selected MIT Data base records are represented in the table 1 with their rhythm types contents for the aim of performance evaluation. Sixteen different patients have been considered in the experiments, to provide at least different arrhythmia cases (N, PVC, RBBB and LBBB). Normal beat (N); Premature ventricular contraction (PVC); Right bundle branch block (RBBB) and Left bundle branch block (LBBB) according to the MITDB annotation files.

2.4 Feature Selection

In Table 2 we present the eleven features descriptors that seem to be most important for characterizing the cardiac arrhythmias needed to classification which has been chosen with the help of specialists in cardiology.

Table 1. The name of selected MIT-BIH records with their rhythm types

Record	Rhythm type			
	Normal	PVC	RBBB	RBBB
100	62	0	0	0
103	58	0	0	0
103	58	0	0	0
106	27	36	0	0
109	0	0	0	102
111	0	0	0	41
116	45	0	0	0
118	0	0	32	0
119	50	34	0	0
124	0	0	33	0
200	0	50	0	0
207	0	0	0	47
208	0	131	0	0
209	25	0	0	0
212	5	0	56	0
214	0	50	0	50
215	73	0	0	0
Total	345	301	121	233

2.5 Artificial Metaplasticity Neural Network

ANNs, widely used in pattern classification within medical fields, are biologically inspired distributed parallel processing networks based on the neuron organization and decision-making process of the human brain [11]. In this paper an MLP, the most commonly used feedforward neural networks due to their fast operation, ease of implementation, and smaller training set requirements and reliability in classification problems in pattern recognition applications [12][13] is used with the aim of classifying cardiological patterns.

The concept of biological metaplasticity was defined in 1996 by Abraham [14] and now is widely applied in the fields of biology, neuroscience, physiology, neurology and others [14], [15]. Recently, Ropero-Pelez [15], Andina [16] and Marcano-Cedeño [17] have introduced and modeled the biological property metaplasticity in the field of ANNs, obtaining excellent results. Among the different AMP models tested by the authors, the most efficient one, in terms of learning time and performance, is the approach that connects metaplasticity and Shannon's information theory, which establishes that less frequent patterns carry more information than frequent patterns [18]. This model defines AMP as a learning procedure that produces greater modifications in the synaptic weights with less frequent patterns and fewer modifications with more frequent ones. Biological metaplasticity favors synaptic strengthening for low-level synaptic activity, while the opposite occurs for high level activity. The model is applicable to general ANNs [16],[17], although in this paper it has been implemented for a multilayer perceptron (MLP).

The Backpropagation (BP) algorithm presents some limitations and problems during the MLP training [19]. The artificial metaplasticity on multilayer perceptron algorithm (AMMLP) tries to improve BP algorithm by including a variable learning rate $\eta(x)$ in the training phase affecting the weights in each iteration step based on an estimation of the real distribution of training patterns. That

Table 2. The various descriptors

Attributes	Meaning
Duration P	The width of the P wave.
PR interval	The distance between the beginning of the P wave and the beginning of QRS.
QRS complex	The distance between the beginning of the Q wave and the end of the S wave.
Duration T	The width of the T wave.
ST segment	The distance between the end of the S wave or R and the beginning of the T wave.
QT interval	The distance between the beginning of QRS and the end of the T wave.
RR previous: RRp	The distance between the peak R of the present beat and the peak R of the previous beat.
RR next : RRn	RRn: the distance between the peak R of the present beat and the peak R of the following beat.
RDI (delay of the deflexion)	From the beginning of QRS to the top of the latest wave of positivity R peak.
Beat duration	The distance between the beginning of the P wave and the end of the wave T.
RRp / RRn.	The ratio RRp / RRn

is if s, j, $i \in$ N are the MLP layer, node and input counter respectively, for each W(t) component $\omega_{ij}^{(s)}(t) \in$ R, where W(t)is the weight matrix, we can express the weight reinforcement in each iteration as:

$$\omega_{ij}^{(s)}(t+1) = \omega_{ij}^{(s)}(t) - \eta(x)\frac{\partial E[W(t)]}{\partial \omega_{ij}}$$

$$= \omega_{ij}^{(s)}(t) - \eta\frac{1}{f_X^*}\frac{\partial E[W(t)]}{\partial \omega_{ij}} \quad (1)$$

Being $\eta \in R^+$ a parameter for the learning rate, E[W(t)] the error function to be minimized and f_X^* the following function assuming that the distribution function of patterns is Gaussian [16],[17].

$$f_X^*(x) = \frac{A}{\sqrt{(2\pi)^N}.e^{B\sum_{i=1}^{N} x_i^2}} \quad (2)$$

Where N is the number of neurons in the MLP input layer, and parameters A and B $\in R^+$ are algorithm optimization values empirically determined which depend on the specific application of the AMLP algorithm. f_X^* has high values for infrequent x values and close to 1 for the frequent ones and can therefore be straightforwardly applied in weights updating procedure to model the biological metaplasticity during learning [17].

3 Results

In this section we present the results obtained in this research. All the models used in this study were trained and tested with the same data and validated

using 10-fold cross-validation. The MLP and AMMLP proposed as classifiers for cardiac arrhythmias were implemented in MATLAB (software MATLAB version 7.4, R2007a) and computer Pentium IV of 3.4 GHz with 2 GB of RAM. The eleven attributes detailed in Table 2 were used as the inputs of the ANNs. Table 3, shows the network structure, metaplasticity parameters, epochs, mean square error (MSE) and numbers of patterns used in the training and the testing of the MLP and AMMLP classifiers.

3.1 The AMMLP Algorithm

1. Network structure used in the experiments:
   (a) Number of input neurons equal to the number of attributes of the records in the database (plus the bias input).
   (b) Number of hidden layers: 1.
   (c) hidden neurons: 8 (a compromise solution found empirically to achieve the results with a simple structure)
   (d) Output neurons: 4 (all classifications present four classes)
   (e) Learning rate η =1
   (f) Activation function is sigmoidal with value between [0,1].
2. Initialize all weights in weight matrix W randomly between [- 1,1]
3. Training phase
   (a) Test training conditions
      i. if epochs = 500
         stop training
      ii. if Mean Squared Error, MSE = 0.01
         stop training

3.2 Network Structure Selection

To select the best configuration for each model used in this study, we tested different network structures and parameters. Table 3 shows the best architectures for each model.

Table 3. Network parameters applying to the MIT-BIH dataset

Types Classifiers	Network Structure			MSE	Epochs	Metaplasti. Parameters		Number of Patterns
	I	HL	O			A	B	
AMMLP	11	8	4	0.001	500	39	0.5	1000
MLP	11	8	4	0.001	500	NA*	NA*	1000

* NA: not apply

3.3 Measures of Quality

To evaluate the performance of the proposed classifier, three measures are used and defined as follows:

$$Sensitivity(SE) = \frac{TP}{TP + FN}(\%) \tag{3}$$

$$Specifity(SP) = \frac{TN}{TN + FP}(\%) \tag{4}$$

$$Accuracy(AC) = \frac{TP + TN}{TP + TN + FP + FN}(\%) \tag{5}$$

Where TP, TN, FP, and FN stand for true positive, true negative, false positive and false negative, respectively. If for example a segment with the V arrhythmia is classified as the V, then it is said that the segment is classified TP. On the other hand if a non-V segment is classified as non-V, then it is said that the segment is classified TN. Any non-V segment which is classified a V segment by mistake will produce a FP, while any V segment which is classified a non-V segment by mistake will produce a FN result.

3.4 Model Evaluation

For test results to be more valuable, a *k-fold* cross-validation is used among the researchers because it minimizes the bias associated with the random sampling of the training [20]. In this method, the whole data are randomly divided into k mutually exclusive and approximately equal size subsets. The classification algorithm is trained and tested k times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus k different test results exist for each training-test configuration [21]. The average of these results provides the test accuracy of the algorithm [20]. A 10-fold cross-validation is used in all of our experiments by separating the selected 1000 samples randomly into 10 subsets with 100 records each and then taking each subset as test data in turns.

3.5 Performance Evaluation

In this study, the models were evaluated based on the accuracy measures discussed above (classification accuracy, sensitivity and specificity). The results were achieved using 10-fold cross-validation for each model, and are based on the average results obtained from the test data set for each fold. The results obtained are showed in Table 4.

As seen in Table 4, the results obtained by AMMLP algorithm are superior to the ones obtained by MLP. The reported results (Table 4), validated by means of the 10-fold cross-validation method, show that the Artificial metaplasticity model produces a higher accuracy than the MLP model. Average of

Table 4. Results for 10-fold cross-validation for all folds and AMMLP and MLP models. Bold values highlight the best results obtained in this research.

Fold *N*	MLP (%)			AMMLP (%)		
	SP	SE	AC	SP	SE	AC
1	98.34	87.37	92.84	98.94	98.87	98.91
2	94.73	98.31	97.06	100	99.43	99.63
3	98.31	87.42	94.51	97.89	100	99.26
4	91.67	91.12	91.18	96.56	98.52	97.59
5	91.27	90.11	90.27	92.73	91.6	92.81
6	96.78	95.65	95.89	97.59	100	99.06
7	100	81.48	89.13	100	96.29	97.83
8	92.47	100	97.12	100	100	100
9	100	93.8	96.89	99.02	100	99.97
10	90.29	98.23	92.34	95.2	98.29	97.47
Average	95.38	92.34	93.72	**97.79**	**98.3**	**98.25**

the AMMLP model was 98.25% accurate with sensitivity and specificity rates of 98.3% and 97.79%, respectively. MLP model obtained a prediction average accuracy of 93.72% with a sensitivity rate of 92.34% and a specificity rate of 95.38%.

For comparison purposes, Table 5 gives the classification accuracies of our method and previous methods applied to the same database. As can be seen from the results, our AMMLP method obtains excellent classification accuracy. We report that the empirical results of AMMLP show a great potential, in terms of improving learning and therefore performance in most cases, no matter what multidisciplinary application it is applied to [21-22].

Table 5. Classification accuracies obtained with our method and other classifiers from the literature

Authors (year)	Method	Accuracy (%)
Hu Y.et al., [22] (1997)	Expert Approach	94.00
Minami K. et al., [23] (1999)	Fourier-NN	98.00
Osowski S et al., [24] (2001)	Fuzzy Hybrid NN	96.06
Owis M.I. et al., [25] (2002)	Blind Source Separation	96.79
Prasad G. et al., [26] (2003)	ANN	85.04
Dayong G. et al., [4] (2004)	NN Bayesian	90.00
Yu S.N. et al., [6] (2008)	ICA-NN	98.71
Benchaib Y. et al., [27] (2009)	MLP BPA	95.12
Gothwal H. et al., [5] (2011)	Fourier-NN	98.48
in this study (2013)	**AMMLP**	**98.25**

4 Discussion

The results obtained by the proposed AMMLP algorithm in this paper are among the best compared with the other state-of-the-art methods. The AMMLP is only beat by Gothwal H. et al., [5] and Yu S.N et al., [6] who obtained accuracies of 98.48% and 98.71% respectively (see Table 5). It must be taken into account that in these two better methods a preprocessing phase is applied, Fast Fourier Transform in one case and Independent Component Analisys in the other. We

believe that if we combine feature selection methods with AMMLP algorithm, our results could significantly improve, because AMMLP algorithm is clearly superior to the MLP used without preprocessing.

5 Conclusions

In this study, the artificial metaplasticity on MLP has been applied to the problem of cardiac arrhythmias classification. The AMMLP approach is based on the biological property of metaplasticity. The goal of this research was to compare the accuracy of the proposed AMMLP with the classical MLP with Backpropagation and also with other state-of-the-art classifiers applied to the MIT-BIH Database. Proposed AMMLP algorithm provides better results than MLP with backpropagation algorithm and is among the best of the state-of-the-art algorithms applied to the same database. The results indicate that the use of the AMMLP algorithm is an alternative option for cardiac arrhythmias detection and could be used as a computer aided detection system for second opinion by physicians when making their diagnostic decisions.

References

1. World Health Organization. Cardiovascular diseases, http://www.euro.who.int/en/what-we-do/health-topics/noncommunicable-diseases/cardiovascular-diseases/definition
2. Haiying, Z.: A New System Dedicated to Real-time Cardiac Arrhythmias Teleassistance and Monitoring. Journal of Universal Computer Science 12(1), 30–44 (2006)
3. Jadhav, S.M., Nalbalwar, S.L., Ghatol, A.A.: Artificial Neural Network Models based cardiac Arrhythmia Disease Diagnosis from ECG Signal Data. International Journal of computer Apllications 44(15), 8–13 (2012)
4. Dayong, G., Madden, M., Chambers, D., Lyons, G.: A Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study. In: Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4 (2005)
5. Gothwal, H., Kedawat, S., Kumar, R.: Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network. Journal of Biomedical Science and Engineering 4, 289–296 (2011)
6. Yu, S.N., Chou, K.T.: Integration of independent component analysis and neural networks for ECG beat classification. Expert Systems with Applications 34(4), 2841–2846 (2008)
7. Holter, N.J.: New methods for heart studies. Science 134, 12–14 (1961)
8. Fogoros, R.N.: The Electrocardiogram (ECG), http://heartdisease.about.com
9. Moody, G.B., Mark, R.G.: The impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medicine and Biology Magazine 20(3), 45–50 (2001)
10. MIT-BIH arrhythmia databse, Harvard-MIT Division of Health Science Technology, Biomedical Health Centre, 1st edn., Cambridge, MA, USA, pp. 1975–1979 (1980)

11. Andina, D., Pham, D.: Computational Intelligence for Engineering and Manufacturing. Springer, The Nederlands (2007)
12. Haykin, S.: Neural Networks a Comprehensive Foundation. MacMillan College Publishing Company, New York (1995)
13. Hagan, M.T., Demuth, H.B., Beale, M.: Neural network design. Thomson Learning, Stamford (1996)
14. Abraham, W.C.: Activity-dependent regulation of synaptic plasticity (metaplasticity) in the hippocampus. In: The Hippocampus: Functions and Clinical Relevance, pp. 15–26. Elsevier Science, Amsterdam (1996)
15. Kinto, E.A., Del Moral Hernandez, E., Marcano Cedeño, A., Ropero-Peláez, J.: A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 628–636. Springer, Heidelberg (2007)
16. Andina, D., Alvarez-Vellisco, A., Jevtic, A., Fombellida, J.: Artificial metaplasticity can improve artificial neural network learning. Intelligent Automation and Soft Computing; Special Issue in Signal Processing and Soft Computing 15(4), 681–694 (2009)
17. Marcano-Cedeño, A., Quintanilla-Dominguez, J., Andina, D.: Breast cancer classification applying artificial metaplasticity algorithm. Neurocomputing 74(8), 1243–1250 (2011)
18. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27, 379–423 (1948)
19. Leung, H., Haykin, S.: The complex backpropagation algorithm. IEEE Transactions on Signal Processing 39(9), 2101–2104 (1991)
20. Frawley, W.J., Paitetsky-Shapiro, G., Matheus, C.J.: From data mining to knowledge discovery: An overview. In: Advances in Knowledge Discovery and Data Mining, pp. 611–620. AAAI Press/The MIT Press (1996)
21. Polat, K., Sahan, S., Gnes, S.: A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. Expert Systems with Applications 31(2), 264–269 (2006)
22. Hu, Y.H., Palreddy, S., Tompkins, W.J.: A patient- adaptable ECG beat classifier using a mixture of experts approach. IEEE Transactions on Biomedical Engineering 44(9), 891–900 (1997)
23. Minami, K., Nakajima, H., Toyoshima, T.: Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. IEEE Transactions on Biomedical Engineering 46(2), 179–185 (1999)
24. Osowski, S., Lin, T.H.: ECG beat recognition using fuzzy hybrid neural network. IEEE Transactions on Biomedical Engineering 48(11), 1265–1271 (2001)
25. Owis, M.I., Youssef, A.B.M., Kadah, Y.M.: Characterization of ECG signals based on blind source separation. Medical and Biological Engineering and Computing 40(5), 557–564 (2002)
26. Prasad, G.K., Sahambi, J.S.: Classification of ECG arrhythmias using multiresolution analysis and neural networks. In: Proceedings of TENCON 2003 IEEE Conference on Convergent Technologies, vol. 1, pp. 227–231 (2003)
27. Benchaib, Y., Chikh, M.: A Specialized learning for neural classification of cardiac arrhythmias. Journal of Theoretical and Applied Information Technology 6(1), 81–89 (2009)

SBRS: Bridging the Gap between Biomedical Research and Clinical Practice

Santiago Timón-Reina, Rafael Martínez-Tomás, Mariano Rincón-Zamorano, Tomás García-Sáiz, Estela Díaz-López, and R.M. Molina-Ruíz

UNED, Hospital Clínico San Carlos

Abstract. The field of Biomedical research is currently one with the greatest social impact and publication volume, providing continuous advances and results which should, to a great extent, reach the general clinical practice. Similarly, direct clinical experience may offer experimental results and conclusions which may lead, guide and foster new investigations. However, this interaction between research and clinical practice is yet too far from being optimal. On one side, research results are published without standardization, suffering terminological issues, which prevent its automatic handling and great scale information treatment/management. On the other, for the practitioner, the task of reviewing papers, bibliography, experimental results, etc. in order to keep updated his everyday clinical practice, is very time consuming, causing not to be done continuously.

The implantation of Information Technologies in the biomedical research field has developed numerous search and bibliographic management resources, existing a current trend towards building and publishing open access terminologies, ontological knowledge models and big datasets with biomedical content. All together, beside Semantic Web technologies, methodologies and Linked Open Data and AI techniques, conforms a technological framework which gives the opportunity to bridge the gap between research and clinical practice to support the physician in evidence based decision making.

In this work, as a starting point to the final aim of linking research and clinical practice, we describe a Semantic Bibliographical Recommender System (SBRS) based on patient profile integrated with electronic health record (EHR) which, without closing the loop, offers to the medical professional the latest and most significant experimental evidences related to his concrete case study. The system's functionality and utility is exemplified through real life psychiatric cases, assisted by an expert psychiatrist.

1 Introduction

The field of Biomedical research is one of most social impacting and publishing volume fields. This fact is easily observable through some bibliographic search systems indexation evolution. To exemplify, we have used "Medline Trend" tool [1] to obtain the evolution of the number of PubMed indexed papers per year,

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 191–200, 2013.

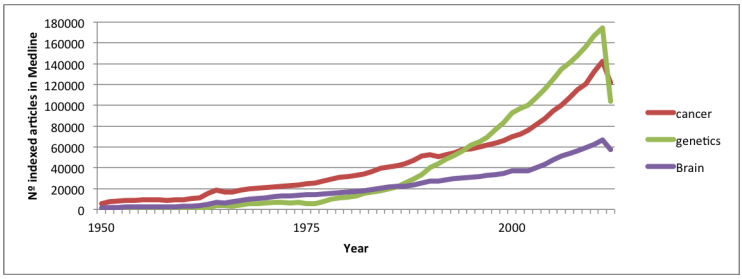


Fig. 1. Publication volume evolution

performing three search queries by general and popular terms in the Biomedical context: cancer, genetics and brain (for the term neuroscience is relatively new). Figure 1 shows this evolution.

The discoveries made by this research field usually have an application in clinical practice, either directly (clinical research) or indirectly (basic science), updating diagnosis, protocols, etc. In the same way, clinical experience may give conclusions and experimental results which impact directly on current research, guiding and fostering new investigations. A very clear example is novel drug tracking in clinical practice, which have its own standards and protocols [2].

However, the interaction between Biomedical research and clinical practice is still far from being efficient. Despite the final aim of research is finding ways of application, divulgation means are specific and closed: few search engines and online journals with few application interfaces. This situation requires a proactive attitude from interested people and agencies looking for published articles. Moreover, research papers lack of the standardization required to allow automatic treatment beyond its indexation. All together results in a very time consuming task, making it difficult for the doctor to remain updated in his daily clinical practice. It is therefore necessary a search for means and specific strategies to ease and encourage implementation of research based recommendations and to ensure changes in practice [3].

Fortunately, with the gradual (and faster everyday) implantation of Information Technologies and Artificial Intelligence techniques, tools have emerged to help the dissemination, search and access to literature, such as PubMed[1]. Likewise, working groups focused on formalizing the terminology along different biomedical knowledge domains started to appear, developing terminologies such as SNOMED CT [4] or MeSH[2] and, a step beyond the terminologies, comprehensive ontology based knowledge models like Gene Ontology (GO) [5] or the Foundational Model of Anatomy (FMA) [6].

[1] http://www.ncbi.nlm.nih.gov/pubmed/
[2] http://www.nlm.nih.gov/mesh/

Following standards and methodologies of the Semantic Web[3] (OWL 2[4], RDF[5], SWRL[6] and SPARQL[7]), these models are enabling the build of great linked, opened and semantic datasets, following Linked Open Data[8] recommendations.

Altogether, we have a) bibliographic search engines focused in Biomedical research, b) controlled terminologies and knowledge models over various biomedical knowledge domains and c) great datasets, giving the opportunity to create, on the one hand, standards for structuring and sharing Biomedical research articles and, on the other hand, tools for exploiting those standards to ease the access to research content for the practitioners.

Given the number and diversity of research domains and the high level of domain-specific expertise that is needed for such an undertaking, it is only reasonable that each discipline take responsibility for developing its research abstraction scheme. However, given the Semantic standards, it will be desirable that there would be an abstraction framework to easily manage disparate research works.

In the field of neuroimaging, the BrainMap database project has been one of the first efforts aiming at the standardization problem [7]. It defines a formal metadata coding scheme to describe the content of functional neuroimaging research, allowing to search across coordinate and brain report locations.

As we have stated, the automatic bibliographic management is a very ambitious goal, requiring the approach and resolution of various problems. The first of them is the low degree of standardization in natural language paper structure and the need for a digital content description beyond terminological indexation.

Despite the necessary degree of standardization is not yet available, taking this technological framework as a starting point, it is possible to design and build systems which may take advantage of available technology. In this work we have developed a prototype of a Semantic Bibliographic Recommender System (SBRS) for the field of neuroimaging which, based on a Patient Profile, generates bibliographic references related to patients characteristics. In order to achieve this goal, we have mapped and extended various bio-ontological resources, generating an ontological model to represent patients and neuroimaging articles. With a system like the one proposed, a practitioner could receive research feedback while seeing patients.

2 Methods

In Web context, user experience has been notably improved in the last years with the introduction and growth of Recommendation Systems. These systems

[3] http://www.w3.org/standards/semanticweb/
[4] http://www.w3.org/TR/2012/REC-owl2-overview-20121211/
[5] http://www.w3.org/standards/techs/rdf#w3c_all
[6] http://www.w3.org/Submission/SWRL/
[7] http://www.w3.org/standards/techs/sparql#w3c_all
[8] http://www.w3.org/standards/semanticweb/data

proactively suggest items which may be of particular interest to the user, based on his behavior and/or preferences[8].

Our initial proposal to improve biomedical research integration within everyday clinical practice is a Semantic Bibliographic Recommender System (SBRS), which follows a Content Based approach [9]. In this kind of systems, items are described based on its features. Since these items are research papers and we are following a semantic approach, these features are concepts from the knowledge model.

In order to build recommendations, the system needs some kind of input or query and a known user profile. This profile usually consists in two types of information:

1. A model of the user's preferences. One common representation is a function which for any item predicts the likelihood that the user is interested in that item. In our case is a semantic description of the patient.
2. A history of the user's interactions with the system. This history helps to improve system's performance by learning and adjusting function's components.

Unlike common Recommender Systems, where the user profile is built on one entity (the user), our scenario requires to split the information between the profile which provides the features of interest and the interaction history. The patient characteristics conform the features of interest, building the Patient Profile (PP), and the interaction history would be built on physician's item choice. At this moment, the system stores user feedback and interaction, but the learning process is not yet implemented.

2.1 Patient and Item Modeling

In order to build such a Recommendation System, we need to define an item description model describing the items the system will have to serve. In our case, the items are neuroimaging research papers (NRP). As we have already seen, Brainmap's metadata codification scheme gives us a good starting point but, since we are looking for semantic interoperability between different domains (anatomy, psychology, genetics, etc.), we need to build a semantic model of this scheme i.e, an ontology.

Fortunately there already exists an ontology covering part of Braimap's coding scheme: the Cognitive Paradigm Ontology [10]. This project intends to formalize, starting from Brainmap Scheme, certain characteristics of the cognitive paradigms used in the fMRI and PET literature. However, it does not fully map concepts with other domain ontologies (such as FMA) and neither every Brainmap metadata is represented. Hence, in order to cover our needs, we have needed to add some extensions, generating the extended CogPOe.

Primarily, the extensions we need are Patient/Subject related. Since our system makes use of patient profiles, the needs can be covered by aligning Patient and EHR oriented ontologies. The Computer-based Patient Ontology (CPR) [11]

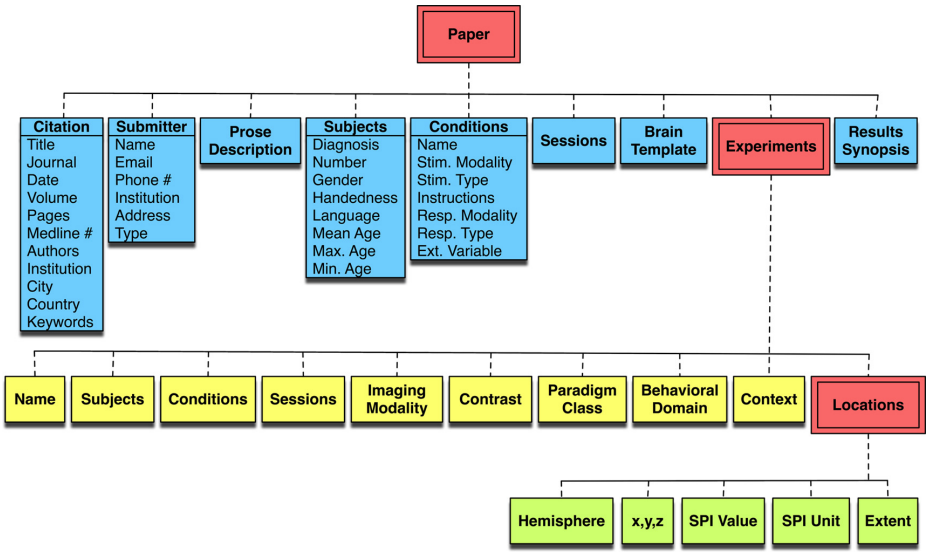


Fig. 2. Brainmap meta-data codification scheme

attempts to define a minimal set of terms that provide grounded, ontologically commitment for the representations shared between many of the healthcare information (such as HL7 RIM)[9], process and terminological models via the use of foundational ontologies.

Briefly, the alignment has been done by referencing CPR concepts from CogPOe. The most important mapping is the relation between Brainmap's Subjects concept, because it defines the subjects involved in an experiment concept and will, indirectly, carry many relevant information. It is represented with cpr:Patient concept. Related patient features (properties), such as diagnostic, gender, medication, etc are also obtained from CPR.

2.2 Building the Patient Profile

As we already said, the system will be based on patient features, which means that it must be build from some sort of EHR data.

But user related data in the context of EHR is a delicate issue because of the existence of segmented EHR systems managing disparate patient representations and, more importantly, the use of various standards among different regulatory agencies.

The best way to deal with this problem is enabling semantic interoperability between EHR systems and regulatory bodies , by implementing a Semantic Mediation System capable of automatically map equivalent fields or concepts between different standards [12]. This is the solution raised by the SALUS project [13,14].

[9] `http://www.hl7.org/implement/standards/rim.cfm`

The Patient Profile building process is designed following the same principles, creating a semantic representation of patient profile from EHR's input.

Therefore, we need to choose an EHR standard which will serve as the system's input. There exist multiple EHR standards so, based on its widespread implementation, we chose HL7's (Health Level 7)[10] Clinical Document Architecture (CDA)[11] were chosen. Briefly, this standard enables documents to be expressed both as free text and coded format, using SNOMED CT as its terminology.

As shown in figure 3, using CDA and SNOMED codes is possible to map physical exploration document concepts to CPR.

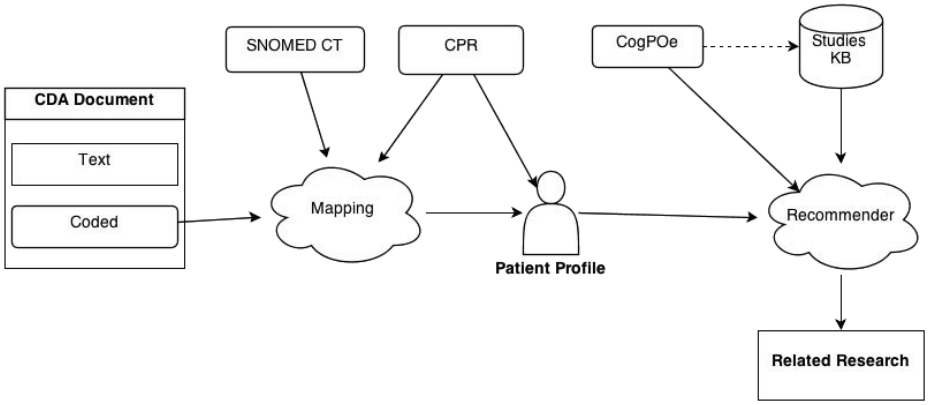


Fig. 3. System overview

2.3 NPR Selection

Both the Patient Profile and item descriptions are instances of our patient centered ontological model, so, in order to evaluate the relevance of a given paper, we have to compare the patient's features with research subject's features. This is achieved with a similarity function which quantifies the similarity between two given instances.

Looking into the ontology mapping literature, we find many similarity measuring methods: hierarchical, graph-based, information theory-based, etc. Since we are looking into instances rather than concepts (i.e., classes) we need to explore the way these instances express its meaning: the implicit labeled graph structure of semantics. We have implemented the similarity function by traversing the ontology instances [15].

According to this method, in order to evaluate the similarity of a pair of given instances, $InstA$ and $InstB$, we need to look into the set of properties which connects them to other elements. Let this instances have properties

[10] `http://www.hl7.org/`
[11] `http://hl7book.net/index.php?title=CDA`

$A = \{property_1, \ldots, property_n\}$ and $B = \{property_1, \ldots, property_m\}$ respectively. For elements connected by a common property i, the similarity measure is computed by the $PropertySM_i$ function.

Then, the similarity measure for two instances is computed as the sum of similarity measures obtained for each property:

$$similarity(InstA, InstB) = \frac{\sum_{i=1}^{|A \cap B|} PropertySM_i(elemtA, elementB)}{|A \cup B|}$$

where $elementA$ and $elementB$ are elements (instances or literals) connected to the evaluated instances $InstA$ and $InstB$ by the i^{th} property respectively. This function gives a similarity measure from the range $[0, 1]$.

At this point, the similarity function sets the same weight factor for each property. This factor is needed to personalize and to determine the importance of a given property, which is context dependent. For example, in our problem, the property hasDisease may have more weight than hasName, since the diagnosis is more important than the name to compute the Patient Profile similarity.

3 Use Case

In order to illustrate the viability of our solution, we have tested our system reproducing, under the proposed structure, with a real clinical case with the collaboration and feedback from an expert psychiatrist focused on Eating Disorder (ED) research.

Classifications of ED (DSM-IV and ICD10) are still mainly focused on the preoccupation for body weight and distortion of body shape, which do not really seem to tell enough about these disorders, specially regarding treatment. More recently, affect dysregulation in ED has been emphasized [16,17], and there is quite a lot of evidence that the emotional awareness and emotion regulation are affected in ED [18]. Most of these ideas were supported by studies that emphasized the relationship between alexithymia and emotional awareness in anorexia. Recent studies are focused in the neurocircuits behind these processes and even recent efforts to reclassify these disorders have suggested classifications of ED within personality subtypes in a: 1) dysregulated/undercontrolled pattern, 2) a constricted/overcontrolled pattern, 3) high-functioning/perfectionistic pattern [19].

Despite the ongoing research efforts, many of these suggestions and discoveries are unnoticed in everyday clinical practice.

As we already stated, at this moment, there is not available any knowledge/semantic database with full-structured research papers. For this reason, we have needed to build a knowledge base using Virtuoso-Opensource[12] triple store where a set of 113 papers, following our ontological model (CogPOe), have been stored. With this dataset we have been able to test the system with the clinical case.

[12] `http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/`

This clinical case consisted in a patient diagnosed with Anorexia Nervosa. The psychopathological exploration identifies that the patient is calmed with collaborative attitude and fluid and coherent speech., but with hypercontrol and obsessive personality traits.

We have tested the system with an example of ED coded in a CDA compliant document serving as the input.From this document, the system maps and builds the PP to compute the set of nearest bibliographic recommendations. The output of this very first query was satisfactory, since 8 of 10 from retrieved NPRs were considered as related to the clinical case, being the 80% of recommendations. Because of the limited dataset size we still have not been able to plan a rigorous performance evaluation (like Precision and Recall), being the dataset population as one of the future works. However, the expert remarked that having this information available from concepts related to the spectrum of these disorders, e.g. ALEXITHYMIA, one of the main characteristics described in AN, would be very useful when trying to decide the best therapy for these patients. For example, patients more focused on emotional regulation should incise in mindfulness therapy. So, this practical example shows us the utility of these integrated tools for the practitioner.

4 Conclusions and Future Work

In this work we have noted a very known problem: the gap between biomedical research and clinical practice. We also have highlighted that this gap is getting narrower as technology is being implanted in Health Care Systems. However, more technology means different systems interoperating and, many of them, overlapping, arising the need for a greater standardization.

As a sample of this standardization, we have created a semantic model starting from Brainmap's metadata coding scheme, reusing, mapping and extending disparate domain ontologies (CogPO and CPR). With this representation of neuroimaging papers, we have been able to propose a Semantic Bibliographic Recommender System to help the practitioner in the bibliographic gathering task. Planning the integration with EHR systems, we have design our system based on the widely used HL7 CDA standard, creating a mapping system to obtain a Patient Profile as the Recommender input.

To test our proposal, we have worked with an expert psychiatrist, receiving advice and feedback. The tests results are promising, encouraging us to keep populating the dataset, refine and scale the system to greater goals.

The concept and the architecture of the Web is naturally evolving to a distributed service oriented scenario, emerging applications which implement interfaces to allow third party applications to interact with them. We think bioinformatic applications should be compliant with this philosophy, therefore, our system is built as a RESTful service, allowing any application to interact with it.

The semantic Patient Profile opens the door to integrate large amounts of clinical data, beyond research bibliography. Linked Open Data initiative is spreading

along many of the Biomedical domains and, with tools like Bioportal and the Datahub[13], it is possible to easily locate, use and share existing biomedical data.

Developing this integration between EHR systems and federated biomedical datasets is no longer a wish, but a fact with on going efforts like the Clinica Mayo's [20].

In the future, we look forward to integrate drug and genetic datasets to automatically enrich the Patient Profile, giving the chance to offer the most complete information for the practitioner.

References

1. Corlan, A.D.: Medline trend: automated yearly statistics of PubMed results for any query (2004), http://dan.corlan.net/medline-trend.html
2. C. for medicinal products for human use (CHMP), ICH guideline E2B (R3) Electronic transmission of individual case safety reports (ICSRs) - implementation guide - data elements and message specification (2011), http://www.emea.europa.eu/docs/enGB/documentlibrary/Scientificguideline/2009/09/WC500002767.pdf
3. Bero, L.A., Grilli, R., Grimshaw, J.M., Harvey, E., Oxman, A.D., Thomson, M.A.: Getting research findings into practice: Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. BMJ 317(7156), 465–468 (1998), http://www.bmj.com/content/317/7156/465
4. Stearns, M., Price, C.: SNOMED clinical terms: overview of the development process and project status. In: Proceedings of the AMIA (2001), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243297/
5. Ashburner, M., Ball, C., Blake, J.: Gene Ontology: tool for the unification of biology. Nature ... (2000), http://www.ncbi.nlm.nih.gov/pmc/articles/pmc3037419/
6. Rosse, C., Mejino, J.L.V.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. Journal of biomedical informatics 36(6), 478–500 (2003), http://www.ncbi.nlm.nih.gov/pubmed/14759820
7. Fox, P.T., Lancaster, J.L.: Opinion: Mapping context and content: the BrainMap model. Nature Reviews Neuroscience 3(4), 319–321 (2002), http://www.ncbi.nlm.nih.gov/pubmed/11967563
8. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User Modeling and User-Adapted Interaction 22(4-5), 317–355 (2012), http://www.springerlink.com/index/10.1007/s11257-011-9115-7
9. Pazzani, M., Billsus, D.: Content-based recommendation systems. In: The adaptive web, pp. 325–341 (2007), http://www.springerlink.com/index/QQ35WT68L6774261.pdf
10. Turner, J.A., Laird, A.R.: The cognitive paradigm ontology: design and application. Neuroinformatics 10(1), 57–66 (2012), http://www.ncbi.nlm.nih.gov/pubmed/21643732
11. W3C, Computer-based Patient Record Ontology (2009), http://code.google.com/p/cpr-ontology/

[13] http://datahub.io/

12. Martínez-Costa, C., de Andrade, A., Karlsson, D.: Towards the harmonization of Clinical Information and Terminologies by Formal Representation. In: EJBI, vol. 8(3) (2012), http://www.ejbi.org/img/ejbi/2012/3/Costa_en.pdf
13. Declerck, G., Hussain, S., Parès, Y., Daniel, C.: Semantic-sensitive extraction of EHR data to support adverse drug event reporting (2011), http://www.srdc.com.tr/projects/salus/docs/ICSR_Reporting_Tool_SWAT4LS_V2.pdf
14. Laleci, G., Yuksel, M., Dogac, A.: Providing Semantic Interoperability between Clinical Care and Clinical Research Domains. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society c, 1–14 (2012), http://www.ncbi.nlm.nih.gov/pubmed/23008263
15. Andrejko, A., Bieliková, M.: Comparing instances of ontological concepts for personalized recommendation in large information spaces. Computing and Informatics 28, 429–452 (2012), http://www.cai.sk/ojs/index.php/cai/article/viewArticle/45
16. Gilboa-Schechtman, E., Avnon, L.: Emotional processing in eating disorders: specific impairment or general distress related deficiency? Depression and ... (2006), http://onlinelibrary.wiley.com/doi/10.1002/da.20163/abstract
17. Hatch, A., Madden, S., Kohn, M.: Anorexia nervosa: towards an integrative neuroscience model. European Eating ... (2010), http://onlinelibrary.wiley.com/doi/10.1002/erv.974/abstract
18. Ortego Saenz de Cabezón, A., Espina Eizaguirre, A.: Alexithymia and its relationships with anxiety and depression in eating disorders. Personality and ... (2004), http://www.sciencedirect.com/science/article/pii/S0191886903000990
19. Thompson-Brenner, H., Westen, D.: Personality subtypes in eating disorders: validation of a classification in a naturalistic sample. The British Journal of Psychiatry (2005), http://bjp.rcpsych.org/content/186/6/516.short
20. Pathak, J., Kiefer, R., Chute, C.: Applying linked data principles to represent patient's electronic health records at Mayo clinic: a case report. In: ... Symposium on International Health ..., pp. 455–464 (2012), http://dl.acm.org/citation.cfm?id=2110415

Perceptual Analysis of Speech Signals from People with Parkinson's Disease

J.R. Orozco-Arroyave[1,2], J.D. Arias-Londoño[1], J.F. Vargas-Bonilla[1], and Elmar Nöth[2]

[1] Universidad de Antioquia, Medellín
[2] Friedrich Alexander Universität, Erlangen-Nürnberg

Abstract. Parkinson's disease (PD) is a neurodegenerative disorder of the nervous central system and it affects the limbs motor control and the communication skills of the patients. The evolution of the disease can get to the point of affecting the intelligibility of the patient's speech.

The treatments of the PD are mainly focused on improving limb symptoms and their impact on speech production is still unclear. Considering the impact of the PD in the intelligibility of the patients, this paper explores the discrimination capability of different perceptual features in the task of automatic classification of speech signals from people with Parkinson's disease (PPD) and healthy controls (HC). The experiments presented in this paper are performed considering the five Spanish vowels uttered by 20 PPD and 20 HC.

The considered set of features includes linear prediction coefficients (LPC), linear prediction cepstral Coefficients (LPCC), Mel-frequency cepstral coefficients (MFCC), perceptual linear prediction coefficients (PLP) and two versions of the relative spectra coefficients (RASTA).

Accordin the results for vowels /e/ and /o/ it is not enough to consider one kind of perceptual features, it is required to perform combination of different coefficients such as PLP, MFCC and RASTA. For the case of the remaining vowels, the best results are obtained considering only one kind of perceptual features, PLP for vowel /a/ and MFCC for vowels /i/ and /u/.

Keywords: Perceptual analysis, Parkinson's disease, linear prediction, relative spectra analysis.

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that results from the death of dopaminergic cells in the substantia nigra, a region of the mid-brain. PD is the second more prevalent neurological disorder after the Alzheimer's disease [1]. About 1% of the people older than 65 suffer from this disease and in Colombia the prevalence of PD is around 172.4 per each 100.000 inhabitants [2]. People with Parkinson's disease (PPD) commonly develop speech impairments affecting different aspects such as respiration, phonation, articulation and prosody [3]. It is already demonstrated that the phonation problems in PPD are related to the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 201–211, 2013.

vocal fold bowing and incomplete vocal fold closure [4], [5]. This behavior also generates intelligibility problems for the speech of the patients, affecting their communications skills and their capability for interacting with other people [6].

Medical treatments such as neuropharmacological and neurosurgical are focused on improving limb symptoms, but their impact on speech production is still unclear [7]. Due to this fact, in the last few years several published works have been focused on the automatic classification of speech recordings from PPD and from healthy controls (HC). In [8] the authors employ acoustic and prosodic features along with features derived from a two-mass model of the vocal folds. For the acoustic modeling of the speech signals, the authors consider a speech recognition model based on Gaussian Mixture Models (GMM) with 13 Mel-frequency Cepstral Coefficients (MFCC) and for the prosodic analysis they include fundamental frequency (F_0), energy, voiced and unvoiced segments and pitch periods all calculated on the voiced segments of a running speech recordings. The reported recognition rates are 88% with the acoustic model (MFCC) and 90.5% with the prosodic features.

In [9], the authors perform the automatic classification of PPD and HC considering four features: Harmonics to Noise Ratio (HNR), Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA) and Pitch Period Entropy (PPE). Their results indicate that, considering this set of features, it is possible to achieve classification rates of up to 91.4%.

On the other hand, the evolution of the PD through the time has been also studied. In [10] the authors form a features set composed by different dysphonia measures and analyze their correlation with the evolution of the Unified Parkinson Disease Rating Scale (UPDRS) in a period of six months. According to their results, the UPDRS scale can be mapped with a precision of up to 6 points, which is very close to the clinician's observations.

Considering that the PPD exhibit a loss of intelligibility in their speech [11], our hypothesis is that including perceptual information in the speech modeling processes is possible to achieve good results in the automatic classification of speech from PPD and HC. The perceptual analysis of pathological speakers have been addressed typically using different kind of coefficients. Some of them have been already used for speech signals from PD [8], but there is still a lack of understanding about the discrimination capabilities that these kind of features can provide for the automatic assessment of speech signals from PPD. Bearing this in mind, this work studies the contribution of six well known representation coefficients estimated over the five Spanish vowels and tries to establish which features are more suitable for the automatic classification of speech from PPD and HC.

The performed experiments include Linear Prediction Coefficients (LPC) [12], Linear Prediction Cepstral Coefficients (LPCC) [13], Mel-frequency Cepstral Coefficients (MFCC) [14], Perceptual Linear Prediction coefficients (PLP) [15] and two versions of the Relative Spectra coefficients (RASTA), those with cepstral filtering (RASTA-PLP-CEPS) and those without cepstral filtering (RASTA-PLP-SPEC) [16].

The rest of the paper is organized as follows. The section 2 includes the description of the methodology tha is being proposed in this work. In the section 3 gives the details of the experiments that are presented in the paper. The section 4 presents the results obtained on each experiment and finally, the section 5 exposed the conclusions derived from the presented work.

2 Methodology

Figure 1 depicts a block diagram of the steps carried out in the methodology presented in this work. The right side of the figure illustrates each stage of the methodology and the left side of the figure shows a brief explanation of that stages. The voice signal is first divided and windowed (with Hamming windows) into frames to perform a short-time analysis. After, the characterization stage takes place. In this work, six different sets of perceptual features are considered: LPC, LPCC, PLP MFCC and two versions of the RASTA-PLP coefficients, one with cepstral filtering and other one without such filtering. Once the features are calculated, a features selection process is performed for each set of features as in [17].

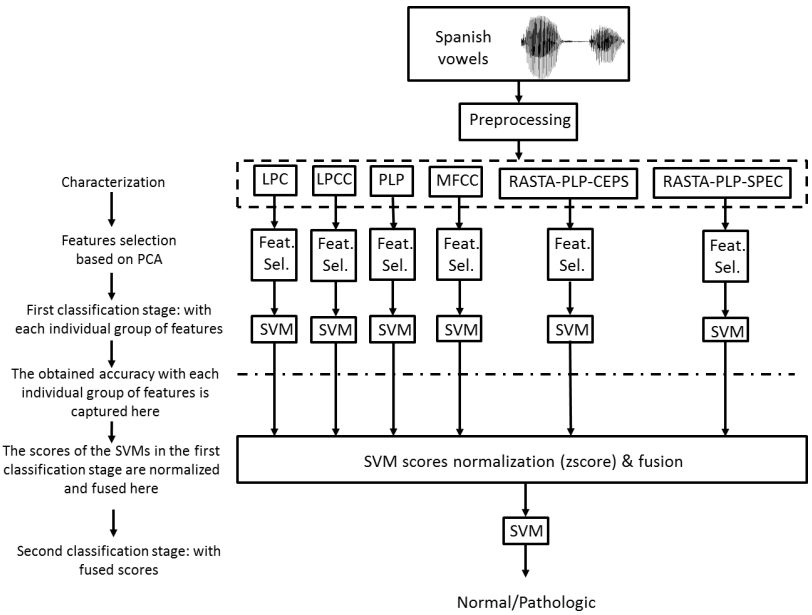


Fig. 1. Methodology for the perceptual analysis of speech from PPD

After the features selection a two layer classification scheme is used. The first stage of the classification process is performed per each kind of perceptual features by means of a support vector machine (SVM) with a radial basis Gaussian kernel. After, the results obtained with each SVM are combined into a new feature space. Once that results are combined, they are normalized using a zscore strategy and finally, the last stage of the classification process is done with another SVM.

In the following subsections, some details of each part of the methodology are presented.

2.1 Perceptual Modeling

As it was pointed out, in this work different perceptual coefficients are implemented with the aim of establish which of them are the more appropriate for the automatic classification of speech from PPD and HC. Considering their wide usage in speech modeling, the LPCs coefficients are included in this work. Linear predictive techniques have been applied not only for the LPCs calculation but also for the formants estimation. The information provided by the LPCs allows to perform articulation analysis in the speech of PPD [18]. This coefficients are able to model the vocal tract as a filter, thus considering that PPD are unable to have a total motor control of their vocal tract, the frequency response of that filter will be also abnormally changed.

A similar modeling to those that is performed by the LPCs in the frequency domain can be made in the cepstral domain. Such analysis is made through the LPCC coefficients. The LPCCs have demonstrated to be more robust in several speech modeling tasks mainly oriented to speech recognition processes [13]. In this work we want to validate their robustness in the automatic classification of speech from PPD and HC.

Additionally the MFCC coefficients have been included in the experiments presented here. This kind of coefficients can model irregular movements in the vocal tract [14] and have demonstrated to be efficient for modeling pathological speech signals, not only in the case of dysphonia detection [14], [19] but also for the analysis of dysarthric speech signals [8].

On the other hand, with the aim of including perceptual information to the modeling that is performed with LPCs, we decided to include the PLP coefficients estimated as in [15]. The LPC analysis assumes the same number of resonances on every frequency bands; however, there are evidences that demonstrate that beyond about 800Hz, the spectral resolution of hearing decreases with frequency [15]. Considering this drawback of the LPC analysis, Hynek Hermansky proposed to use a critical-band filtering over the linear predictive analysis in order to improve the resolution of the analyzed bands.

Other kind of features that are included in this work are the RASTA-PLP coefficients. This technique was presented by Hermansky and Morgan in [16].

The main assumption of this modeling strategy is that the human perception is less sensitive to slowly varying stimuli, thus it is possible to make the speech analysis less sensitive to slowly changing factors. For doing that, the critical-band filter that is used for PLP modeling is replaced by a filter bank with a sharp spectral zero at the zero frequency. The RASTA-PLP coefficients can be represented in spectral or in cepstral domains and the difference between both approaches is that in the spectral domain the resulting set of coefficients is according to the number of filtered bands and in the cepstral domain the resulting set of coefficients is according to the number of cepstral coefficients. In this work we use both representations to analyze their influence in the process of automatic classification of speech from PPD and HC.

2.2 Features Selection and Classification

The features selection strategy that is used in this work is based on principal component analysis and it was implemented as in [17]. After the application of the features selection algorithm the system has a sub set of features that is optimal in terms of its variance content.

The classification between recordings from PPD and HC is performed with a SVM that is trained with a Gaussian kernel [20]. This classifier is used because of its extensive usage in the state of the art for the automatic classification of pathological and healthy voices [9], [8], [17].

3 Experiments

3.1 Recording and Corpus of Speakers

The data for this study consists of speech recordings from 20 PPD and 20 HC sampled at $44.100Hz$ with 16 quantization bits. All of the recordings were captured in a sound proof booth. The people that participated in the recording sessions are balanced by gender and age: the ages of the men patients ranged from 56 to 70 (mean 62.9 ± 6.39) and the ages of the women patients ranged from 57 to 75 (mean 64.6 ± 5.62). For the case of the healthy people, the ages of men ranged from 51 to 68 (mean 62.6 ± 5.48) and the ages of the women ranged from 57 to 75 (mean 64.8 ± 5.65). All of the PPD have been diagnosed by neurologist experts and none of the people in the HC group has history of symptoms related to Parkinson's disease or any other kind of movement disorder syndrome.

The recordings consist of sustained utterances of the five Spanish vowels, every person repeated three times the five vowels, thus in total the database is composed of 60 recordings per vowel on each class. This database is built by *Universidad de Antioquia* in Medellín, Colombia.

3.2 Experimental Setup

The experiments performed in this work are carried out following the methodology exposed in figure 1. The voice recordings are preprocessed by Hamming windows of 40ms with and overlap of 20ms. The perceptual features are calculated for each windowed frame of speech. Each voice signal is represented by the sets of feature vectors, each vector contains the values of the parameters for each frame. The final feature vector per voice signal is composed by the estimations of mean, standard deviation, skewness and kurtosis of the values obtained per feature trough the frames.

The number of coefficients is fixed in 12 for every kind of perceptual features but for the case of RASTA-PLP-SPEC a total of 27 coefficients are estimated because this is the number of frequency bands that are filtered. For RASTA-PLP-CEPS the number of coefficients is 12 because it is the number of cepstral coefficients used. Considering that four statistics are taken from each kind of features, the sets of parameters contain a total of 48 measures for the case of LPC, LPCC, PLP, MFCC and RASTA-PLP-CEPS. For the case of RASTA-PLP-SPEC the number of measures is 108.

The tests performed over the proposed system have been made following the strategy indicated in [21]. The 70% of the data are used for the feature selection and for training the classifier and the remaining 30% is for testing; the different subsets for training and testing are randomly formed. As it was exposed in section 2, for each pair of training and testing subsets the two stages of classification are made: the first is when only each set of perceptual coefficients are considered individually and the second is when the scores obtained in the first classification stage are combined. Each stage of the classification process is repeated ten times per each pair of subsets (training and testing), forming a total of 100 independent realizations of the experiment.

In order to look for the best performance of the system, the scores obtained in the first classification stage are incrementally combined. The order of that fusion is according to the classification rate obtained in the individual classification stage.

4 Results and Discussion

The results obtained in the first stage of the classification process, when each subset of perceptual features are used per each vowel, are presented in table 1. The highlighted items correspond to the best results obtained over all of experiments. Note that for the vowels /a/, /i/ and /u/ the best results correspond to those obtained with only one kind of perceptual coefficients. For vowel /a/ the PLP parameters exhibited the best results, while for vowels /i/ and /u/ the best features were the MFCCs in both cases. The obtained results for the

Table 1. Results obtained in the automatic classification of speech signals from PPD and HC using each kind of perceptual features individually

Vowel	Features	Individual accuracy	Specificity	Sensitivity
/a/	LPC	60,58 % 6,98	65,44 % 12,20	55,72 % 13,97
	LPCC	64,86 % 8,08	60,83 % 14,10	68,89 % 7,49
	MFCC	59,72 % 9,04	66,33 % 15,95	53,11 % 7,60
	PLP	**76,19 % 9,04**	**80,22 % 10,69**	**72,17 % 11,05**
	RASTA-PLP-CEPS	56,72 % 6,17	41,22 % 8,98	72,22 % 7,33
	RASTA-PLP-SPEC	59,47 % 7,58	60,11 % 12,31	58,83 % 10,47
/e/	LPC	66,25 % 6,82	61,56 % 12,85	70,94 % 9,53
	LPCC	63,41 % 8,45	60,17 % 14,37	66,67 % 12,48
	MFCC	66,97 % 11,98	64,61 % 16,46	69,33 % 14,87
	PLP	66,39 % 7,23	67,17 % 9,57	65,61 % 13,93
	RASTA-PLP-CEPS	66,08 % 6,04	64,67 % 9,03	67,50 % 10,60
	RASTA-PLP-SPEC	71,28 % 7,38	66,06 % 13,14	76,50 % 9,51
/i/	LPC	58,17 % 8,76	53,72 % 12,14	62,61 % 14,83
	LPCC	71,61 % 6,9	71,44 % 10,58	71,78 % 11,32
	MFCC	**75,30 % 8,43**	**78,78 % 10,20**	**71,83 % 10,53**
	PLP	70,83 % 7,49	70,50 % 7,07	71,17 % 12,10
	RASTA-PLP-CEPS	66,33 % 5,39	59,94 % 14,71	72,72 % 12,97
	RASTA-PLP-SPEC	69,64 % 5,59	67,44 % 6,59	71,83 % 8,22
/o/	LPC	59,22 % 8,42	64,89 % 12,30	53,56 % 11,55
	LPCC	71,83 % 5,88	82,78 % 6,62	60,89 % 8,39
	MFCC	78,31 % 5,32	87,17 % 8,72	69,44 % 9,39
	PLP	69,97 % 6,84	71,06 % 12,88	68,89 % 9,65
	RASTA-PLP-CEPS	71,11 % 6,89	70,22 % 19,27	72,00 % 15,90
	RASTA-PLP-SPEC	68,61 % 7,62	61,17 % 7,70	76,06 % 10,24
/u/	LPC	62,61 % 8,37	63,94 % 9,31	61,28 % 8,19
	LPCC	64,86 % 7,67	62,44 % 10,36	67,28 % 10,84
	MFCC	**76,28 % 6,11**	**82,44 % 10,75**	**70,11 % 7,14**
	PLP	73,14 % 11,11	76,89 % 9,44	69,39 % 13,16
	RASTA-PLP-CEPS	67,78 % 6,75	52,89 % 12,86	82,67 % 9,27
	RASTA-PLP-SPEC	62,94 % 5,89	52,61 % 11,17	73,28 % 10,26

vowels /e/ and /o/ have not any highlighted row due to the best results for that vowels were obtained when several perceptual coefficients are combined. Such results are presented in table 2. The results obtained in the second stage of the classification process are presented in table 2. Note that in this case the best results are obtained for vowels /e/ and /o/ when five subsets of features are combined. For the case of vowel /e/ the considered features are RASTA-PLP-SPEC, MFCC, PLP, LPC and RASTA-PLP-CEPS while for vowel /o/ the set of features include MFCC, LPCC, RASTA-PLP-CEPS, PLP and RASTA-PLP-SPEC.

It is interesting to note that the best results in vowels /e/ and /o/ include MFCC, PLP and both versions of the RASTA-PLP coefficients for both vowels. It indicates that with the aim of achieving better results in the automatic classification of speech signals from PPD and HC, the characterization with perceptual features must consider more than one kind of coefficients for vowels /e/ and /o/. In general, the performance obtained with the vowels /e/ and /o/ are higher than in the other cases; however, note that for reaching such results the inclusion of five features were required.

Table 2. Results obtained when each subset of features are incrementally combined

Vowel		Two subsets	Three subsets	Four subsets	Five subsets	six subsets
	Accuracy	70,67 % 10,63	71,89 % 7,33	70,39 % 6,53	70,06 % 7,58	68,28 % 6,32
/a/	**Specificity**	72,94 % 14,17	78,67 % 11,31	76,67 % 11,70	75,17 % 9,59	66,50 % 70,65
	Sensitivity	68,39 % 11,53	65,00 % 10,75	64,50 % 8,81	64,64 % 8,80	70,06 % 8,67
	Accuracy	73,11 % 7,33	74,11 % 7,67	75,86 % 6,16	**77,22 % 6,28**	77,28 % 7,00
/e/	**Specificity**	62,28 % 13,64	64,39 % 12,47	65,33 % 14,47	**66,39 % 13,78**	67,94 % 14,56
	Sensitivity	81,50 % 10,67	83,56 % 12,44	86,50 % 9,48	**88,06 % 7,87**	86,78 % 6,50
	Accuracy	70,47 % 11,00	72,53 % 8,66	74,86 % 8,78	75,00 % 6,69	74,78 % 5,69
/i/	**Specificity**	68,78 % 11,40	74,50 % 9,29	74,06 % 10,68	69,44 % 11,27	69,00 % 11,71
	Sensitivity	72,06 % 18,14	70,44 % 20,21	75,67 % 18,67	80,56 % 13,86	80,78 % 13,93
	Accuracy	78,92 % 8,89	80,81 % 7,39	79,77 % 6,77	**81,08 % 6,82**	80,78 % 6,91
/o/	**Specificity**	83,44 % 6,48	80,56 % 9,51	81,72 % 8,27	**80,22 % 8,02**	73,39 % 7,74
	Sensitivity	74,39 % 14,56	81,06 % 9,39	77,83 % 7,93	**81,94 % 7,71**	82,11 % 7,73
	Accuracy	76,08 % 10,76	75,64 % 7,97	76,5 % 7,36	76,25 % 6,48	76,14 % 7,13
/u/	**Specificity**	77,33 % 12,80	67,28 % 9,50	68,50 % 8,85	67,22 % 7,22	65,72 % 8,61
	Sensitivity	74,17 % 9,68	83,72 % 8,02	84,56 % 8,36	85,33 % 6,71	86,67 % 7,89

The table 3 shows the best results obtained for each vowel in terms of sensitivity and specificity. The best sensitivity and specificity are obtained with vowels /e/ and /u/ respectively; however, note that the sensitivity and specificity are more balanced for the vowel /o/. In order to show the best results per vowel in a more compact way and following the methodology presented in [21], the figure 2 with the detection error tradeoff (DET) curve is included. Note that the more balanced behavior is exhibited with the vowel /o/, while the vowels /a/, /e/ and /i/ show very unbalanced results.

Table 3. Best results obtained per vowel

Vowels	**/a/**	**/e/**	**/i/**	**/o/**	**/u/**
Accuracy	76,19 ± 9,04	77.22 ± 6,28	75.30 ± 8,43	81.08 ± 6,82	76.28 ± 6,11
Sensitivity	72,17 ± 11,05	88,06 ± 7,87	71,83 ± 10,53	81,94 ± 7,71	70,11 ± 7,14
Specificity	80,22 ± 10,69	66,39 ± 13,78	78,78 ± 10,20	80,72 ± 8,02	82,44 ± 10,75

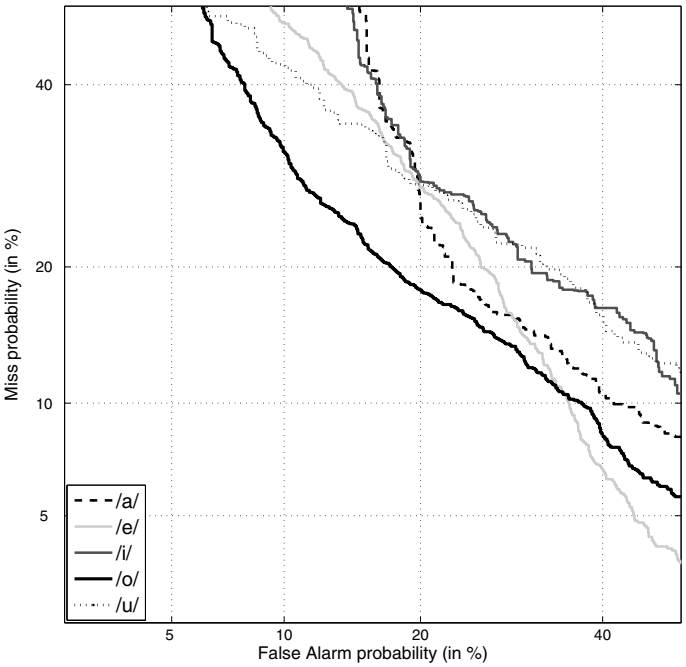


Fig. 2. DET curve for each vowel

5 Conclusions

The perceptual analysis of speech signals from people with Parkinson's disease is performed and the effectiveness, in terms of the discrimination capability, of each kind of perceptual coefficients in the problem of automatic classification of speech from PPD and HC is analyzed.

According to the results, PLP coefficients considered alone, offer good performance just for the case of the vowel /a/ and the MFCCs exhibit the higher performance in the vowels /i/ and /u/. However, for all experiments, the best performance is obtained with the vowel /o/ when MFCC, LPCC, RASTA-PLP-CEPS, PLP and RASTA-PLP-SPEC are merged and considered together.

The RASTA analysis has exposed good results in the task of automatic speech recognition; however, according to our findings, in order to achieve better results with vowels /e/ and /o/, those coefficients must be combined with other perceptual features such as MFCC and LPCC.

The main finding of this work indicates that for vowels /e/ and /o/ it is not enough to consider one kind of perceptual features, it is required to perform combination of different parameters to achieve good results in the task of automatic classification of speech signals from PPD and HC.

Acknowledgments. Juan Rafael Orozco Arroyave is under grants of "Convocatoria 528 para estudios de doctorado en Colombia, generación del bicentenario, 2011" financed by COLCIENCIAS. The authors give a special thanks to all of the patients and collaborators in the foundation "Fundalianza Parkinson-Colombia". Without their valuable support it would be impossible to address this research.

References

1. de Rijk, M.: Prevalence of parkinson's disease in europe: A collaborative study of population-based cohorts. Neurology 54, 21–23 (2000)
2. Sánchez, J., Buriticá, O., Pineda, D., Uribe, C., Palacio, L.: Prevalence of parkinson's disease and parkinsonism in a colombian population using the capture-recapture method. International Journal of Neuroscience 113, 175–182 (2004)
3. Skodda, S., Visser, W., Schlegel, U.: Vowel articulation in parkinson's diease. Journal of Voice 25(4), 467–472 (2011)
4. Hanson, D., Gerratt, B., Ward, P.: Cinegraphic observations of laryngeal function in parkinson's disease. Laryngoscope 94(3), 348–353 (1984)
5. Perez, K., Ramig, L., Smith, M., Dromery, C.: The parkinson larynx: tremor and videostroboscopic findings. Journal of Voice 10(4), 353–361 (1996)
6. Weismer, G., Jeng, Y., Laures, J., Kent, R., Kent, J.: Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatrica et Logopaedica 53, 1–18 (2001)
7. Ramig, L., Fox, C., Shimon, S.: Speech treatment for parkinson's disease. Expert Review Neurotherapeutics 8(2), 297–309 (2008)
8. Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., Rusz, J.: Detection of persons with parkinson's disease by acoustic, vocal and prosodic analysis. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp. 478–483 (2011)
9. Little, M.A., McSharry, P., Hunter, E., Spielman, J., Ramig, L.: Suitability of dysphonia measurements for telemonitoring of parkinson's disease. IEEE Transactions on Bio-Medical Engineering 56(4), 1015–1022 (2009)
10. Tsanas, A., Little, M., McSharry, P., Ramig, L.: Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. IEEE Transactions on Biomedical Engineering 57(4), 884–893 (2010)
11. Falk, T., Chan, W., Shein, F.: Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility. Speech Communication 54(5), 622–631 (2012)
12. Buzo, A., Gray, A., Gray, R., Markel, J.: Speech coding based upon vector quantization. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 15–18 (1980)
13. Kim, H., Choi, S., Lee, H.: On approximating line spectral frequencies to lPC cepstral coefficients. IEEE Transactions on Speech and Audio Processing 8(2), 195–199 (2000)
14. Godino-Llorente, J., Gómez-Vilda, P., Blanco-Velasco, M.: Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. IEEE Transactions on Biomedical Engineering 53(10), 1943–1953 (2006)
15. Hermansky, H.: Perceptual linear predictive (plp) analysis of speech. Journal of the Acoustical Society of America 87(4), 1738–1752 (1990)

16. Hermansky, H., Morgan, N.: Rasta processing of speech. IEEE Transactions on Speech and Audio Processing 2(4), 578–589 (1994)
17. Orozco-Arroyave, J., Vargas-Bonilla, J., Arias-Londoño, J., Murillo-Rendón, S., Castellanos-Domínguez, G., Garcés, J.: Nonlinear dynamics for hypernasality detection in spanish vowels and words. Cognitive Computation 4(2), 1–10 (2012)
18. Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. The Journal of the Acoustical Society of America 129(1), 350–367 (2011)
19. Arias-Londoño, J., Godino-Llorente, J., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Domínguez, G.: An improved method for voice pathology detection by means of a hmm-based feature space transformation. Pattern Recognition 42, 3100–3112 (2010)
20. Scholköpf, B., Smola, A.: Learning with Kernel. The MIT Press (2002)
21. Sáenz-Lechón, N., Godino-Llorente, J., Osma-Ruiz, V., Gómez-Vilda, P.: Methodological issues in the development of automatic systems for voice pathology detection. Biomedical Signal Processing and Control 1, 120–128 (2006)

Characterization of Speech from Amyotrophic Lateral Sclerosis by Neuromorphic Processing

Pedro Gómez-Vilda[1], Ana Rita M. Londral[2], José Manuel Ferrández-Vicente[3], and Victoria Rodellar-Biarge[1]

[1] Grupo de Informática Aplicada al Tratamiento de Señal e Imagen, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Madrid
[2] Instituto de Medicina Molecular, Universidade de Lisboa, Lisboa, Portugal
[3] Dpto. Electrónica, Tecnología de Computadoras, Univ. Politécnica de Cartagena, 30202, Cartagena
pedro@pino.datsi.fi.upm.es

Abstract. Amyotrophic Lateral Sclerosis is a severe disease which dramatically reduces the speech communication skills of the patient as illness progresses. The present study is devoted to define accurate and objective estimates to characterize the loss of communication skills, to help in monitoring illness progress, and in rehabilitating speech by specialists. The methodology proposed is based on the perceptual (neuromorphic) definition of speech dynamics, concentrated in vowel sounds in character and duration. A longitudinal study carried out on an ALS patient during a year is shown, its results discussed and conclusions for a further study are also presented.

1 Introduction

The detection and characterization of vowel spaces is of most importance in many applications, as in pathological characterization of speech, therefore the present work will concentrate in specific vowel representation space detection and characterization by neuromorphic methods. Amyotrophic Lateral Sclerosis (ALS) is a highly impairing neuromotor disease of unclear origin [1] which affects severely the capability of muscles to respond to neural activation, resulting in a progressive decay in voluntary and involuntary movements of the patient, and progressively to a state incompatible with life. One of the most dramatic and observable effects is the gradual degeneration of speech production resulting from progressive affection of the complex neuromuscular system involved in respiration, phonation, swallowing and lingual and oro-facial muscle management. This conducts to a specific kind of disarthria characterized by hypernasality, reduced speech rhythm, vowel intelligibility degradation, loss of consonantal dynamics, reduced and prolonged number of inter-phonation intervals and pitch dysprosody. The patient experiences a loss of oral communication capability which eventually may lead to isolation and depression. The present paper is intended to explore some of the most perceivable symptoms among the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 212–224, 2013.

ones described above, with the aim of helping in early detection and establishing a differential diagnose, as well as providing the speech therapist with objective tools to evaluate the regression process and to optimize exercising techniques helping the patient to keep communication skills as more efficient as possible and for a longer time. Among the different numbers of symptoms mentioned above the paper concentrates in the description of vowel colour and count as possible markers of ALS dysarthria, as well on the number and duration of inter-phonation intervals. In doing so, an important definition has to be established, which is the nature and characteristics of vowel sounds. This is not a trivial task as vowels may be defined under the acoustic-phonetic or phonologic point of view [2]. In such task perceptual concepts of vowel production and perception may be of great help. In what follows vowels will be characterized by the following descriptors: phonation must be present (i.e. a glottal source excitation of the vocal tract must be detected), strong and narrow formant descriptors must be evident, stability in the formants has to be maintained under a certain criterion, and a mapping in the vowel triangle of the patient may be attributed to a certain phonologic attractor or vowel representation space. The paper is organized as follows: A neuromorphic description of the phonation and articulation processes is given in section 2 to help understanding the underlying neuromotor mechanisms involved. A brief description of vowel nature based in formant characteristics and dynamics is given in section 3. In section 4 the metrics used to characterize articulation dysarthrias is presented, and the case study is described. Section 5 presents the analysis results, which are illustrated graphically and briefly discussed. Conclusions are presented in section 6.

2 Neuro-physiological Speech and Phonation Model

Speech production is planned and instantiated in the linguistic neuromotor cortex (see Fig. 1). The neuromotor activation sequence involved in speech production is transmitted to the pharynx (2), tongue (3), larynx (4), chest and diaphragm (5) through the subthalamic secondary units. Fine muscular control is provided by a sophisticated feedback control system (6). Neural speech activation patterns are transmitted through the jugular foramen from the hypothalamic system to the glosopharyngeal and vagus nerves (Cranial Nerves CN IX and X) in several derivations innervating the following muscular structures: levator veli palatini, palatoglosus and palatopharyngeous (2), acting on the naso-pharyngeal switch. These structures play a most relevant role in nasalization (hyper-, hypo- and modal). The superior, middle and inferior pharyngeal constrictors, and stylopharyngeous (3) muscles found in the mid-pharynx, are responsible for the swallowing function as well as of changes in the vocal tract during speech articulation. The cricothyroid, transverse and oblique arytenoid, as well as the posterior cricoarytenoid (4) muscles in the larynx are responsible for vocal fold stretching, adduction and abduction by acting on the cricoarytenoid joint as well as in raising and lowering the cricothyroid cartilage. The vagus nerve (5) is responsible for filling and depleting the lung cavity with air

by contraction and relaxation of the crural diaphragm. Most of the muscles in the tongue, responsible of articulation gestures are innervated by the hypoglosal nerve (CN XII). Other muscle-nerve systems of interest in articulation are the buccal and mental nerves, derived from the maxilary (V_2 branch of the trigeminal) and mandibular nerves (V_3 branch of the trigeminal) responsible of upper and lower lip activity. Any alteration in the functionality of these structures will produce perturbations in the respiration, phonation and articulation giving place to specific dysarthrias [3] [4] which may be characterized by the F_2 vs F_1 positions in time [2].

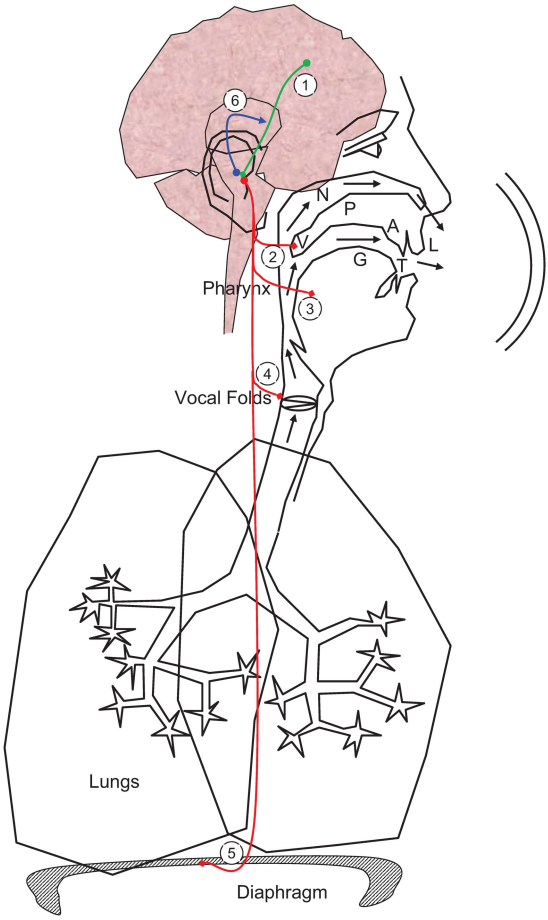


Fig. 1. Simplified view of main neural pathways involved in the production of phonation and speech articulation. N: nasal cavity, V: velum, P: palate, A: alveoli, L: lips, T: teeth, G: tongue.

3 Neuromorphic Characterization of Speech

Vowels may be formally defined as applications between the space of acoustic representations at the cortical level to the set of perceptual symbols defined as vowels at the phonologic or linguistic level [2]. The acoustic-phonetic nature of these patterns is based on the association of the two first resonances of the Vocal Tract, which are referred to as 'formants', and described as F_1 and F_2. F_1 in the range of 200-800 Hz is the lowest, F_2 sweeps a wider range, from 500 to 3000 Hz. Under this point of view the nature of vowels may be described by formant stability during a time interval larger than 30 ms, and relative position in the F_2 vs F_1 space, in which is called the 'Vowel Triangle' (see Fig. 2). The characterization of vowels by neuromorphic speech processing requires the identification of formants as the basic instantiations to develop further knowledge. Formants are specific resonances of the articulation organs (vocal and nasal tracts, and pharyngeal cavities). Neuromorphic processing refers to processing methods directly inspired in neuronal activity (Hebbian structures) [5] [6]. The main processes mimicking the neuronal activity are lateral inhibition formant profiling, tonotopic frequency band tracking, vowel representation space grouping by space-frequency neuromorphic density functions, vowel assignment by mutual exclusion, and vowel temporal clipping, among others.

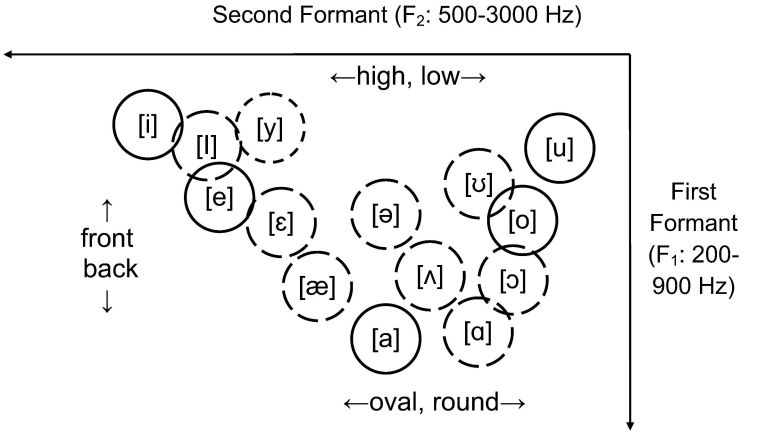


Fig. 2. Subset of the Reference Vowel Triangle for the present study. The vowel set i, e, a, o, u is referred as the cardinal set. Depending on the phonologic set (representation spaces) used by a language other acoustic realizations (dash line) may be assigned to nearby phonologic representations. For instance [æ] in Spanish could be perceptually assigned to /a/.

A full description of these processes can be found in [2]. The specific procedures implemented in this study are the following:

1. The speech trace is processed by an adaptive lattice gradient filter to obtain the inverse vocal tract transfer function.
2. A spectrogram is evaluated from the coefficients of the inverse vocal tract transfer function.
3. Formants are estimated from the maxima of the spectrogram by lateral inhibition.
4. Speech activity and phonation activity are estimated by the umbralization of speech and glottal residual energy.
5. The first two formants are tracked using space-frequency density functions.
6. Vowel representation spaces are used in detecting vowel presence by formant pair associations.
7. Vowel assignments by mutual exclusion determine the most probable vowel uttered.
8. Vowel limits in time are delimited by characteristic-frequency overlapping.

Speech may be described as a time-running acoustic succession of events (or phonetic sequence, see Fig. 3.a) [7]. Each event is associated with an oversimplified phonation paradigm composed of vowels, and non-vowels. Non-vowel sounds are characterized by unstable formants (dynamic), by not having a representation inside the vowel triangle, or by lacking a neat F_2 vs F_1 pattern. The International Phonetic Alphabet (IPA) [8] has been used, with symbols between square brackets [a] and bars /a/ are phonemes (acoustic representations) and phonologic representations, respectively. Formants are characterized in this spectrogram (middle template) by darker energy envelope peaks. What can be observed in the figure is that the vowels and vowel-like sounds correspond to stable positions of the formants.

4 Materials and Methods

The present study has a marked exploratory nature. Early work in formant descriptions for ALS induced dysarthrias has a long history [9]. Nevertheless, objective characterization of these articulation deficiencies using objective representations on the vowel triangle are not frequently found. The study will concentrate in producing sequences of positions on the vowel triangle F_2 vs F_1 corresponding to pairs $F_1(n)$, $F_2(n)$, where n is the discrete time index, as given in Fig. 3 (c and d). The more stable a vowel will be the more points will be found in a given area of the vowel triangle in time. A measurement of the vowel triangle extension covered by a given sentence or utterance may be produced in terms of the distribution of pairs F_1, F_2 in that specific area. Therefore the following landmarks of the vowel triangle will be defined.

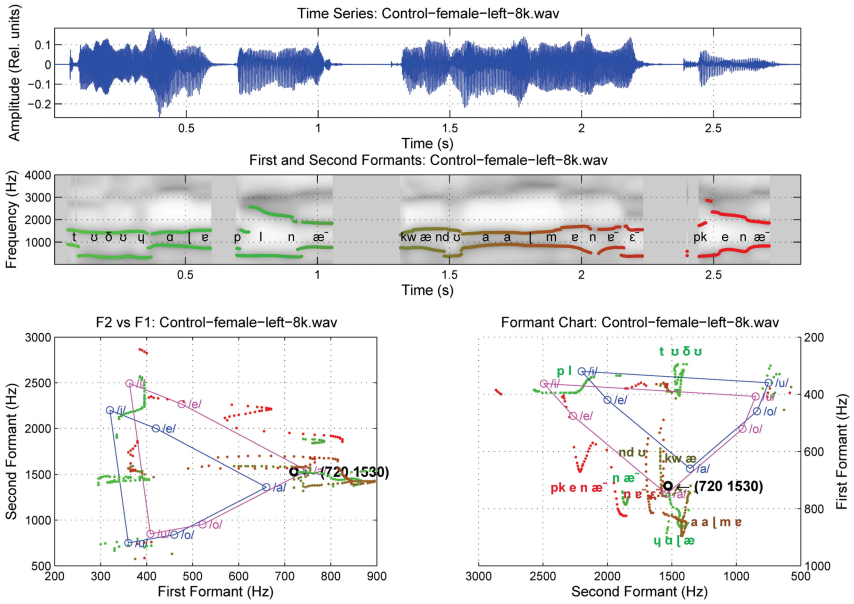


Fig. 3. Fig. 3 a) time series of the utterance in Portuguese "tudo vale a pena quando a alma no pequena" uttered by a female control speaker. b) Adaptive Lineal Prediction Spectrogram (grey background) and first two formants (superimposed in color). The color dots mark the positions of each pair F_1,F_2 from green (the oldest) to red (the most recent). The approximate phonetic labeling is given as a reference. c) Formant plot of F_2 vs F_1. d) Same plot as a Formant Chart commonly used in Linguistics. The blue triangle and circles give the limit positions of the five cardinal vowels /i/, /e/, /a/, /o/, /u/ (for a typical male speaker in blue, female in melba). These plots show the formant trajectories of the utterance. There is color correspondence between the bottom and middle templates to track formant trajectories on the time axis.

$$\begin{aligned} V_{UL} &= \left\{ q_1^{\theta_1}, q_2^{\theta_3} \right\} \\ V_{LL} &= \left\{ q_1^{\theta_1}, q_2^{\theta_1} \right\} \\ V_{MR} &= \left\{ q_1^{\theta_3}, q_2^{\theta_2} \right\} \\ C_{MM} &= \left\{ q_1^{\theta_2}, q_2^{\theta_2} \right\} \end{aligned} \tag{1}$$

where V_{UL}, V_{LL}, V_{MR} and V_{MM} are respectively the uper left, lower left and mid right vertices of the vowel triangle, and C_{MM} is the median centre of the triangle, defined in terms of the generic quantiles

$$q_i^{\theta} = \arg \left\{ \frac{\int_{q_i^{\theta}}^{\infty} \gamma(\nu_i) d\nu}{\int_{-\infty}^{\infty} \gamma(\nu_i) d\nu} < \theta \right\} \tag{2}$$

where $\gamma_i(\nu)$ is the probability distribution of the formant i in frequency ν, and θ is the specific quantile threshold (for instance $\theta = 0.03$ would correspond to a 3 per cent quantile). In the present study the following definitions apply: $\theta_1 = 0.03, \theta_2 = 0.5$ and $\theta_3 = 0.97$. Using these definitions the virtual centroid of the vowel triangle would be defined as

$$C_\nu = \left\{ \frac{q_1^{\theta_1} + q_1^{\theta_3}}{2}, \frac{q_2^{\theta_1} + q_2^{\theta_3}}{2} \right\} \tag{3}$$

whereas the median centroid of the vowel triangle would be

$$C_m = \left\{ q_1^{\theta_2}, q_1^{\theta_2} \right\} \tag{4}$$

The asymmetry coefficient would be the difference between the median and the virtual centroids, which may be expressed in module and argument as

$$M_A) = \left[\left(\frac{2q_1^{\theta_2}}{q_1^{\theta_1} + q_1^{\theta_3}} - 1 \right)^2 + \left(\frac{2q_2^{\theta_2}}{q_2^{\theta_1} + q_2^{\theta_3}} - 1 \right)^2 \right]^{1/2}$$

$$\varphi_A = \arctan \left(\frac{2q_2^{\theta_2} - q_2^{\theta_1} + q_2^{\theta_3}}{2q_1^{\theta_2} - q_1^{\theta_1} + q_1^{\theta_3}} \right) \tag{5}$$

Given the exploratory character of the present study a case study from a ALS female patient has been used in contrast with a control healthy female subject. The case study consisted in four recordings from the female patient taken at specific 3-month intervals, these being referred to as HA-T0 (November 2011), HA-T1 (January 2012), HA-T3 (July 2012) and HA-T4 (October 2012). In all cases the recordings contained utterances of the sentence "tudo vale a pena quando a alma no pequena" in Portuguese. The results of the study conducted on these recordings are given in the next section.

5 Results and Discussion

The descriptions of the vowel triangle for each utterance produced were obtained and compared. For the sake of brevity only the first and last plots compared against the control subject are given in Fig. 4.

It may be seen that the distribution of the control subject stresses the main patterns and trajectories of the target sentence. The first utterance of the ALS patient (HA-T0) stresses the differences in the vocalic space between [ʊ] and [a], but fails in weighting the respective distribution of each vowel group. The last utterance (HA-T4) is clearly unbalanced towards [a], with most of the vowels improperly articulated as [æ]. The plots in Fig. 5 help in establishing a better comparison among the different vowel triangles and to derive resolving conclusions.

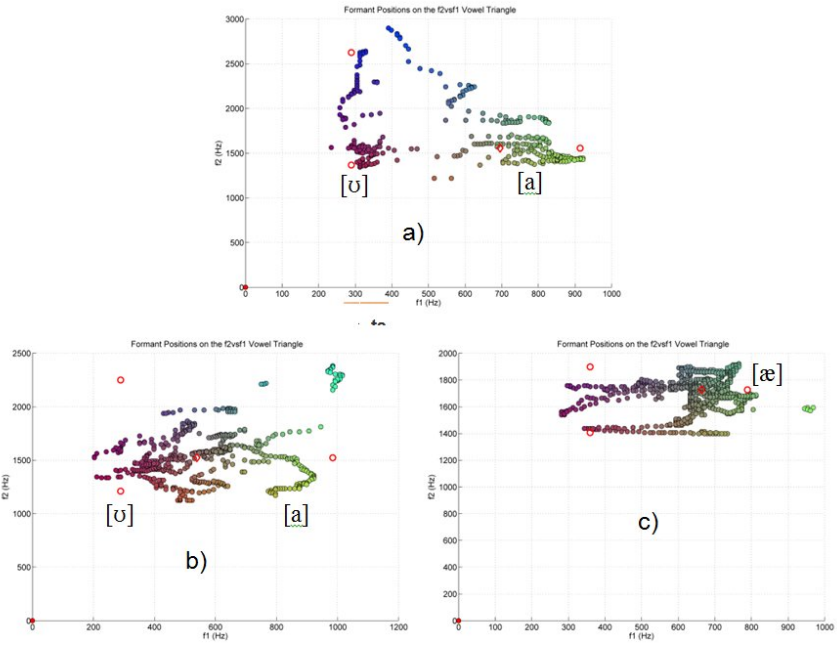


Fig. 4. Positions of f_1 vs f_2 on the vowel triangle. The 3 per cent quartiles are given in red circles. The median centre is given as a red diamond. a) Control female. The main vowel positions are for [ʊ] (in melba) and [a] (in green). The formant trajectories are well organized and separate. The frequency span is large both for f_1 and f_2. b) HA-T0. The vowel space is much more confuse, there are not clear vowel trajectories, but the [ʊ] and [a] are still differentiated . The frequency span is still wide, but it shrinks in f_2. The median centre is slightly tilted to the left. c) HA-T4.

The progressive degradation of the vowel triangle can be clearly perceived, with strong differences between the results for HA-T0 (still comparable with the control subject) and the three last utterances. This would indicate that a strong decay in articulatory ability of the patient took place since November 2011 to January 2012. These results are summarized in Table 1.

The most interesting fact to be stressed is that the angle of the normalized asymmetry coefficient swings from the third to the first quadrant in a progressive succession. This means that the orientation of the vowel median centroids are evolving from a more balanced situation to a tendency marked by vowel [æ]. The two last columns in the table give the Vowel Space Area (VSA) and the Formant Centralization Ratio (FCR) evaluated following [10] [11] as a further

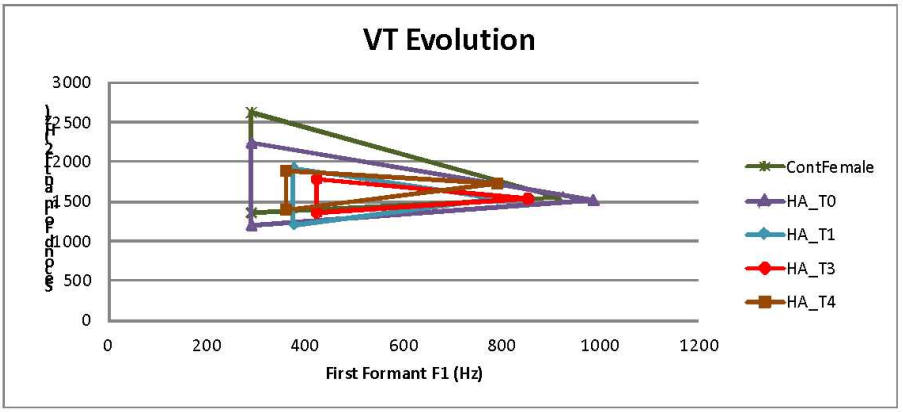


Fig. 5. The approximate vowel triangles derived from the utterances from the control subject (ContFemale), and the ALS patient in four different sessions cronologically ordered from less severity to most severity (HA-T0, HA-T1, HA-T2 and HA-T3) are compared among themselves. It may be seen that the vowel triangle for HA-T0 is slightly narrower in f_2 but larger in f_1 than the control one. HA-T1 shrinks clearly with respect to HA-T0. The same happens with HA-T3 and HA-T4 with respect to HA-T1, although there is not a clear change between themselves.

Table 1. Asymmetry measurements for the vowel triangles studied

.	Cv1(Hz)	Cv2(Hz)	Cm1(Hz)	Cm2(Hz)	M_A	φ_A(deg)	VSA	FCR
MaleRef	445	1578	406	1507	0.099	-152.9	141910	1.15
FemaleRef	570	1828	523	1656	0.125	-131.1	240230	1.07
ContFemale	601	2000	695	1555	0.272	-55.0	456300	0.93
HA-T0	640	1734	539	1523	0.200	-142.3	361240	1.02
HA-T1	586	1578	578	1531	0.033	-113.4	147190	1.29
HA-T3	640	1578	648	1539	0.027	-63.4	90637	1.42
HA-T4	578	1656	664	1726	0.154	15.7	105740	1.43

reference. It may be seen that the FCR gives also an indication of the anomalous articulation function. The respective values of the asymmetry modulus and phase (Relative Center Displacement) are given also in Fig. 6. It may be seen from the plot that the cases show a progression from the third quadrant to the first one, with a transit through the fourth quadrant. This is an objective measurement expressing the migration of the articulation center of gravity towards the position of an open mid vowel as [æ].

The last part of the results presented is the study of vowel and stop interval durations. To produce such results vowels are detected using a coincidence function between the first and second formant CF neuron firings [2], as these

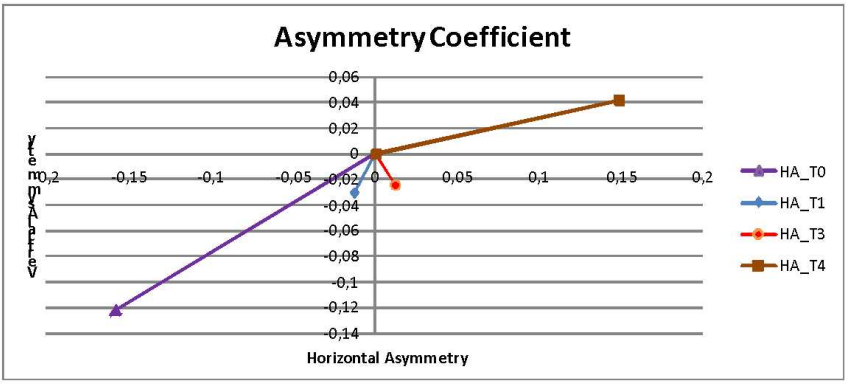


Fig. 6. The asymmetry coefficient is the normalized distance between the triangle base and height midpoint intersection and the median center. This relative displacement is plotted in module and angle for four progressive ALS stages from the same patient (HA-T0, HA-T1, HA-T2 and HA-T3).

Table 2. Vowel group duration and stop and silence intervals

.	**Long Vowels** (L > 150 ms)	**Short Vowels** (50<L<150 ms)	**Short Dyn Groups** (L<50 ms)	**Stops** (50<L<150 ms)	**Silences** (L > 150 ms)
ContFemale	4	3	6	2	4
HA-T0	5	4	12	3	4
HA-T4	6	5	5	1	2

units are activated when a formant is relatively unchanged for a certain interval. Coinciding quasi-stable formants are an indication of vowel presence, independently of its nature. The results for the target sentence from the control and the first and last ALS patient utterances are given in Fig. 7. The presence of vowel groups are divided in three categories: larger than 150 ms, between 50 and 150 ms, and smaller than 50 ms. Classically regular vowels would be included in the second category. The first category may be associated with an impaired articulation. The presence of the third category is associated with dynamic transitions between stable vowel positions, and its reduction can be also associated with impaired articulation. Silences can be also classified according to the same principles. Intervals longer than 150 ms can be associated with phrase splits, under 150 ms with plosive consonants. The account of the different vowel and silence intervals for the three cases presented in Fig. 7 are listed in Table 2.

In general it may be observed that the number of longer vowel groups is larger in pathological speech than in normal speech. It may be seen also that the duration of the utterance is much larger as pathology expresses its severity.

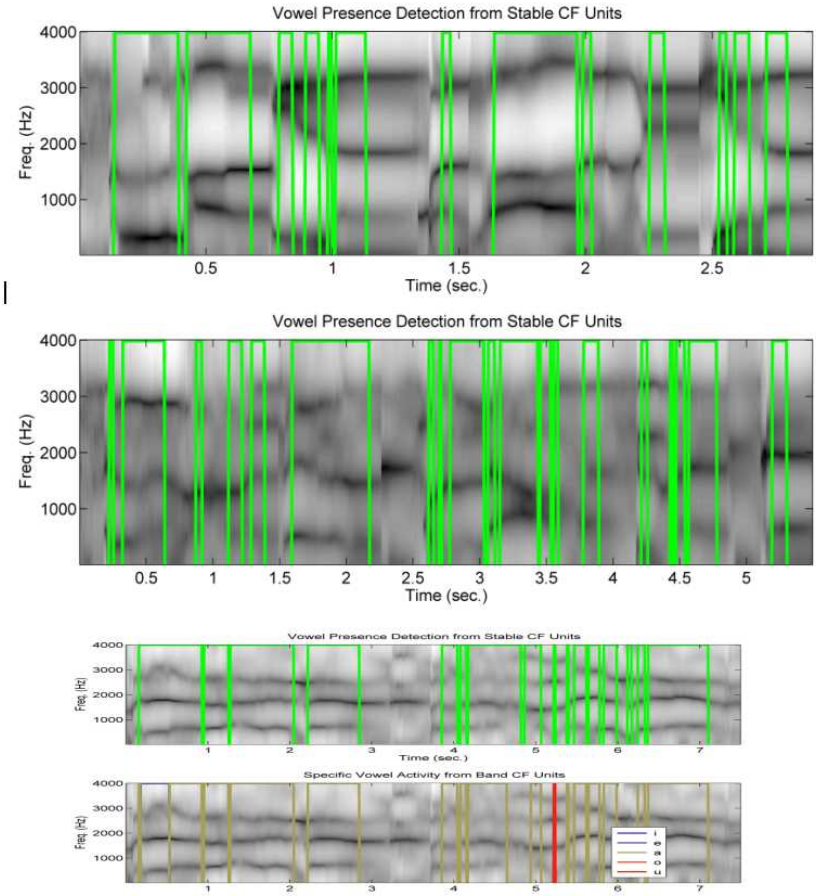


Fig. 7. Vowel-Consonant Dynamics. The green line delimits vowel intervals corresponding to the high level, whereas the low level corresponds to silence or unstable formants (consonant patterns). The long intervals are associated to vowel patterns, the short intervals correspond to brief and unstable vowels near consonantal groups. a) Control female. Stable vowel groups are 200-300 ms long. Short unstable vowels around 50 ms may be observed near plosive groups [p-] and [kw-]. Vowel formants take different configurations. b). HA-T0. The number of long vowel groups is smaller, the number of short vowel groups is larger. Formant patterns deteriorate. c) HA-T4. Larger vowel groups may be appreciated again for the first part of the sentence, but formant patterns are much deteriorated pointing to the positions for [æ]. The number of silences is shorter and larger, the length of the sentence is more than twice longer than in (a).

6 Conclusions

The results of the study avail some of the preliminary goals formulated in section 1, consisting in producing objective measurements of speech degradation phenomena which may be perceived by the expert listener or the speech therapist: increased hypernasality, reduced speech rhythm, vowel intelligibility degradation, loss of consonantal dynamics, reduced and prolonged number of inter-phonation intervals and pitch dysprosody, among others. Due to the limitations of the present study only rhythm, duration and intelligibility of vowels have been studied using neuromorphic detection of formant dynamics to establish the presence and nature of produced vowels. The most important findings established in this sense as illness progresses are the following:

- The utterance of the same sentence is produced in a longer interval.
- The duration of vowels in syllabic nuclei are also extended.
- The number of inter-phrasal intervals is reduced, but its duration is extended.
- The number of pre-stop silence intervals is reduced. As a consequence stop consonants are lost.
- The vowel triangle shrinks, especially in F2.
- The vowel triangle centroids evolve towards [æ].

This last finding needs a further explanation. It is well known from literature that F1 is very much related to the degree of opening of the vocal tract ([i] and [u] corresponding to the more closed extremes, whereas [a] gives the more open extreme), whereas F_2 is more related to the articulation position (where [u] is considered a back vowel whereas [i] is a frontal, and [a] would be a middle vowel). Thus forcing frontal or rear vowels would imply the operation of the hypoglosal and mandibular (mental) neuromotor systems which need not be active in the neutral mid position open vowel given by [æ]. Under severely impaired neuromuscular activity this would be the only articulatory position and the relative colouring of the different vowels would be fused towards this final position. Other vowels nearby the lower vertex of Fig. 2 could also be plausible solutions to an impaired articulatory situation. There are other aspects of ALS dysarthric speech which have not been checked in the present study, as estimating the degree of hypernasality due to the failure of the levator veli palatini, palatoglosus and palatopharyngeous neuromuscular structures acting on the naso-pharyngeal switch, as these would require a spectral detector to model the zeroes in the vocal and nasal tract anti-resonances. Consonantal dynamics could also be traced using neuromorphic speech processing [6]. Dysprosody could also be characterized using well-known pitch tracking methods. These tasks are left for future research. Another important task to be accomplished is the estimation of the biomechanical parameters of phonation in ALS patients, in a similar way used in other neurological pathologies [12], which has not been conducted in this case due to the limitations of the study. Another important task to fulfil in the near future is the collection of a large database containing longitudinal studies as the one described to extend the statistical significance of the findings produced in this study.

Acknowledgements. This work is being funded by grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Science and Technology of Spain.

References

1. Núñez-Batalla, F., Díaz-Molina, J.P., Costales-Marcos, M., Moreno-Galindo, C., Suárez-Nieto, C.: Neurolaryngology. FActa Otorrinol. Esp. 63, 132–140 (2012)
2. Gómez, P., Ferrández, J.M., Rodellar, V., Fernández, R.: Simulating the Phonological Auditory Cortex: From Vowel Representation Spaces to Categories. Neurocomputing (2012) (in press), doi:http://dx.doi.org/10.1016/j.neucom.2012.07.036
3. Yunusova, Y.: Articulatory Movements During Vowels in Speakers With Dysarthria and Healthy Controls. J. Speech, Lang. and Hear. Res. 51, 596–611 (2008)
4. Bongioanni, P.: Communication Impairment in ALS Patients: Assessment and Treatment. In: Maurer, M.: (eds.) Amyotrophic Lateral Sclerosis (2012), http://www.intechopen.com/books/amyotrophic-lateral-sclerosis
5. Gómez, P., Ferrández, J.M., Rodellar, V., Fernández, R.: Time-frequency Representations in Speech Perception. Neurocomputing 72, 820–830 (2009)
6. Gómez, P., Ferrández, J.M., Rodellar, V., Alvarez, A., Mazaira, L.M., Olalla, R., Muñoz, C.: Neuromorphic detection of speech dynamics. Neurocomputing 74(8), 1191–1202 (2011)
7. Greenberg, S., Ainsworth, W.H.: Speech processing in the auditory system: an overview. In: Greenberg, W.A.S. (ed.) Speech Processing in the Auditory System, pp. 1–62. Springer, New York (2004)
8. http://www.arts.gla.ac.uk/IPA/ipachart.html
9. Weismer, G., Martin, R., Kent, R.D., Kent, J.F.: Formant trajectory characteristics of males with amyotrophic lateral sclerosis. J. Acoust. Soc. Am. 91, 1085–1098 (1992)
10. Sapir, S., Ramig, L.O., Spielman, J., Fox, C.: Acoustic Metrics of Vowel Articulation in Parkinson's Disease: Vowel Space Area (VSA) vs. Vowel Articulation Index (VAI). In: Manfredi, C. (ed.) Proc. of MAVEBA 2011, pp. 173–175. Florence University Press (2011)
11. Sapir, S., Ramig, L.O., Fox, C.: Formant Centralization Ratio: A proposal for a New Acoustic Measure of Dysarthric Speech. J. Speech, Lang. and Hear. Res. 53, 114–125 (2010)
12. Gómez-Vilda, P., Rodellar-Biarge, V., Nieto-Lluis, V., Muñoz-Mulas, C., Mazaira-Fernández, L.M., Ramírez-Calvo, C., Fernández-Fernández, M., Toribio-Díaz, E.: Neurological Disease Detection and Monitoring from Voice Production. In: Travieso-González, C.M., Alonso-Hernández, J.B. (eds.) NOLISP 2011. LNCS (LNAI), vol. 7015, pp. 1–8. Springer, Heidelberg (2011)

Fast Approximate Inference in Hybrid Bayesian Networks Using Dynamic Discretisation

Helge Langseth[1], David Marquez[2], and Martin Neil[2]

[1] Department of Computer and Information Science,
The Norwegian University of Science and Technology, Norway
[2] School of Electronic Engineering and Computer Science,
Queen Mary University of London, London E1 4NS, United Kingdom

Abstract. We consider inference in a Bayesian network that can consist of a mix of discrete and continuos variables. It is well known that this is a task that cannot be solved in general using a standard inference algorithms based on the junction-tree. A common solution to this problem is to discretise the continuous variables to obtain a fully discrete model, in which standard inference can be performed. The most efficient discretisation procedure in terms of cost of inference is known as *dynamic discretisation*, and was published by Kozlov and Koller in the late 90's. In this paper we discuss an already published simplification to that algorithm by Neil et al. The simplification is in practise orders of magnitudes faster than Kozlov and Koller's technique, but potentially at the cost of some lack of precision. We consider the mathematical properties of Neil et al.'s algorithm, and challenge it by constructing models that are particularly difficult for that method. Some simple modifications to the core algorithm are proposed, and the empirical results are very promising, indicating that the simplified procedure is feasible also for very challenging problems.

1 Introduction

By a *hybrid* Bayesian network we denote a network where some variables are discrete, others are continuous. It is well known that the exact inference schemes currently employed (e.g., [1]) only work for some particular classes of hybrid BNs. The most common strategies for performing inference in a hybrid BN can roughly be divided into three categories: Firstly, a subset of models (commonly referred to conditional Gaussian models) allow exact inference.Secondly, approximate inference procedures like stochastic sampling can be employed. Finally, one can make explicit changes in the representations of the conditional distribution functions defined for each variable in order to facilitate exact inference. The most prominent among these alternations is *discretisation*, i.e., to "translate" all continuous variables into discrete ones [2]. During discretisation, each new discrete variable has to be given an adequate number of states to capture the associated continuous variable sufficiently well, and the tradeoff between model precision and model complexity is therefore particularly evident during this task.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 225–234, 2013.

Consider a continuous variable with k continuous parents, and assume we discretise each variable into m states. Doing so, we create conditional probability tables with a total size of order $\mathcal{O}(m^k)$ to represent the discretised model. This makes a fine-grained discretisation computationally ineffective, even for moderate values of k. Thus, traditional discretisation heuristics like "equal-mass" and "equal-width" discretisation have been replaced by techniques that use *non-uniform* discretisation. Here, a finer discretisation is employed were it pays the most, an idea pioneered by Kozlov and Koller [3], who aimed at finding a discretised probability density function (pdf) as close to the original pdf as possible in terms of the KL distance [4]. In the following we consider a pdf over a multivariate vector $\boldsymbol{X}$. We assume that $\boldsymbol{X}$ is continuous, as it is trivial to extend our discussion to hybrid domains. Let the pdf $f(\boldsymbol{x})$ be defined on $\boldsymbol{x} \in \boldsymbol{\Omega} \subseteq \mathbb{R}^d$. $\boldsymbol{\Omega}$ is partitioned into hypercubes (or "subsets") $\boldsymbol{\omega}_k$, $k = 1, \ldots, t$ such that $\boldsymbol{\omega}_i \cap \boldsymbol{\omega}_j = \emptyset$ and $\cup_{k=1}^{t} \boldsymbol{\omega}_k = \boldsymbol{\Omega}$. A discretisation $\bar{f}(\boldsymbol{x})$ on $f(\boldsymbol{x})$ wrt. $\{\boldsymbol{\omega}_k\}_{k=1}^{t}$ is a non-negative function of $\boldsymbol{x} \in \boldsymbol{\Omega}$, constant on each hypercube $\boldsymbol{\omega}_\ell$ (i.e., $\bar{f}(\boldsymbol{x}) = \bar{f}_\ell$ for constant $\bar{f}_\ell$ when $\boldsymbol{x} \in \boldsymbol{\omega}_\ell$), and which integrates to unity (so $\sum_{k=1}^{t} \bar{f}_k \cdot |\boldsymbol{\omega}_k| = 1$, where we use $|\boldsymbol{\omega}_\ell| = \int_{\boldsymbol{x} \in \boldsymbol{\omega}_\ell} d\boldsymbol{x}$ to denote the volume of the hypercube $\boldsymbol{\omega}_\ell$). The KL distance from f to $\bar{f}$ can be calculated by summing over the partitions of the discretisation [4,3], giving

$$D\left(f \,\|\, \bar{f}\right) = \sum_{k=1}^{t} \int_{\boldsymbol{x} \in \boldsymbol{\omega}_k} f(\boldsymbol{x}) \log\left(\frac{f(\boldsymbol{x})}{\bar{f}_k}\right) \mathrm{d}\boldsymbol{x}. \tag{1}$$

It is easy to verify that given the subsets $\{\boldsymbol{\omega}_k\}_{k=1}^{t}$, the discretised pdf closest to $f(\boldsymbol{x})$ in KL distance is found by choosing $\bar{f}_\ell = \int_{x \in \boldsymbol{\omega}_\ell} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} / |\boldsymbol{\omega}_\ell|$, see [3]. Finding a "good" discretisation of $f(\boldsymbol{x})$ for $\boldsymbol{x} \in \boldsymbol{\Omega}$ therefore amounts to determining the set of hypercubes $\boldsymbol{\omega}_k$ as defined above.

Define further the notation that $f_\ell^{\uparrow} = \max_{\boldsymbol{x} \in \boldsymbol{\omega}_\ell} f(\boldsymbol{x})$ and $f_\ell^{\downarrow} = \min_{\boldsymbol{x} \in \boldsymbol{\omega}_\ell} f(\boldsymbol{x})$. Now, Kozlov and Koller [3] showed that the contribution from each term in the sum of Equation (1) can be bounded above by

$$\begin{aligned} &\int_{\boldsymbol{x} \in \boldsymbol{\omega}_\ell} f(\boldsymbol{x}) \log\left(\frac{f(\boldsymbol{x})}{\bar{f}_\ell}\right) \mathrm{d}\boldsymbol{x} \leq \\ &\qquad \left[\frac{f_\ell^{\uparrow} - \bar{f}_\ell}{f_\ell^{\uparrow} - f_\ell^{\downarrow}} f_\ell^{\downarrow} \log\left(\frac{f_\ell^{\downarrow}}{\bar{f}_\ell}\right) + \frac{\bar{f}_\ell - f_\ell^{\downarrow}}{f_\ell^{\uparrow} - f_\ell^{\downarrow}} f_\ell^{\uparrow} \log\left(\frac{f_\ell^{\uparrow}}{\bar{f}_\ell}\right)\right] |\boldsymbol{\omega}_\ell| \,. \end{aligned} \tag{2}$$

This motivates a greedy discretisation strategy, where the following two steps are repeated until some termination criteria is met: *i*) Calculate the upper-bound of the contribution to the total KL distance on each hypercube $\boldsymbol{\omega}_k$, $k = 1, \ldots, t$, using Equation (2), and *ii*) Partition the hypercube with the highest bound into two parts.

Kozlov and Koller [3] reported that while this strategy works well for Bayesian networks without evidence inserted, it can become quite poor when posterior probabilities are calculated from (low-probability) evidence. This motivates *dynamic* discretisation, where the discretisation process is conducted while *taking*

```
1  Function dynDiscMarginal
   Input  : BN 𝓑, Evidence ϵ, Query 𝒬.
   Output: Approximation of the posterior distribution f(q|ϵ) in 𝓑.
2  Ω ← Initial discretisation of X;
3  Y ← X \ ϵ;
4  repeat
5  |  𝒟 ← Discretised version of 𝓑 using Ω ;
6  |  Calculate P_𝒟(Y|ϵ);
7  |  foreach variable in Y do
8  |  |  Ω_i ← sort(Ω_i, bestSplit(Ω_i, P_𝒟(Y_i|ϵ)));
9  |  end
10 until converged;
11 return P_𝒟(Q|ϵ);
```

Algorithm 1. Skeleton for the *dynamic discretisation*-algorithm

evidence into account. The discretisation is an integral part of the inference algorithm, and the discretisation is done at the level of the cliques, thus ensuring a close-to-optimal discretisation of the domain as a whole. Unfortunately, though, this algorithm has computational issues, which have prevented it from being commonly used in practice.[1] For instance, the algorithm requires re-implementation of the message passing algorithm for inference (communicating objects called "weights", which are used to re-adjust the discretisation when evidence is found in low-probability regions of the density together with the standard messages), it uses specialised data structures called binary split partition trees for which the standard inference operations must be defined, and it must find or approximate the minimum and maximum values of potentially high-dimensional functions on each hypercube to utilise the bound in Equation (2).

Neil et al. [5] proposed a refinement of Kozlov and Koller's work, defining an algorithm which operates using only the *standard inference engine for discrete variables*. The key idea of the algorithm is to discretise each variable *separately* by utilising Equation (2) to discretise each variable based on that variable' approximated *posterior marginal distribution*. This results in an inference procedure, which is extremely fast, also for BN models of considerable size.[2] The main steps are given in Algorithm 1: The algorithm takes a BN $\mathcal{B}$ over variables $\boldsymbol{X}$, evidence $\boldsymbol{\epsilon}$ and a query Q as input, and starts by roughly discretising all variables $\boldsymbol{X}$. Evidence variables are discretised by defining split-points just below and just above their observed values; these discretisations will not be refined later. Unobserved variables, called $\boldsymbol{Y}$ for ease of reference, are initially discretised by dividing their support into a predefined number of intervals. The discretised version of the

[1] To the best of our knowledge, there is no implementation of Kozlov and Koller's algorithm publicly available.

[2] The dynamic discretisation algorithm [5] is implemented in the AgenaRisk software package. A free version of AgenaRisk can be downloaded from `http://www.agenarisk.com/`

hybrid BN $\mathcal{B}$ is denoted $\mathcal{D}$, and we can use a standard inference engine for discrete Bayesian networks to calculate any posterior distribution in $\mathcal{D}$; $P_{\mathcal{D}}(\boldsymbol{Y}|\boldsymbol{\epsilon})$ is used to explicitly denote that $P(\boldsymbol{Y}|\boldsymbol{\epsilon})$ is calculated in $\mathcal{D}$. The discretisation for variable Y_i (denoted by Ω_i) is refined by repeating the following steps: Firstly, the posterior distribution over $\boldsymbol{Y}$ is calculated in $\mathcal{D}$ (variables defined using the current discretisation). Secondly, in line 8, the discretisation Ω_i for Y_i is refined by adding a new split-point at the mid-point of the current interval with the highest KL bound (calculated using Equation (2)) if this is deemed beneficial; the actual split-point is found by the external function `bestSplit`. The iteration is terminated as soon as a convergence measure is met. In the following we will examine Algorithm 1 in more detail, and discuss the major differences between that algorithm and Kozlov and Koller's work [3].

2 Discretising a Single Variable

The first important difference between [5] and [3] is that Algorithm 1 uses the *marginal discretised pdf* when calculating $f_\ell^{\downarrow}$, $f_\ell^{\uparrow}$ and $\bar{f}_\ell$ which later are used to find the best interval to split (Line 8). Kozlov and Koller, on the other hand, use the (potentially multivariate) continuous functions. Algorithm 1 thus considers the discretised pdf as a step-function (see Fig. 1), where the minimum and maximum values are easily established: $f_\ell^{\downarrow}$ for an interval ω_ℓ is simply defined as the minimum value obtained by the discretised pdf on ω_ℓ and its two neighbouring intervals, and $f_\ell^{\uparrow}$ is found similarly. $\bar{f}_\ell$ is defined as the pdf value at ω_ℓ, although special rules are employed if ω_ℓ holds a mode of the discretised pdf, in which case $\bar{f}_\ell$ is defined as the average value over the three intervals. We note that this approach typically will over-estimate the KL bound in Equation (2), and in particular when the derivative of the true pdf is large in absolute value, thus leading to slightly higher discretisation effort than optimal in those areas. Indications of this effect can be seen in Fig. 2, where Part (a) shows the results of discretising the standard Normal. The optimal discretisation found by simulated annealing is shown with a solid line, and the results of Algorithm 1 are shown with a dashed line. The results are not that different in Part (a), which is based on 10 split-points. However, Part (b) shows the discontinuity points for a discretisation using 24 split-points; the points chosen by Algorithm 1 are in the upper row (drawn as circles) and the approximate optimal solution found by simulated annealing is shown in the lower row (crosses). It is evident that Algorithm 1 puts less effort than optimal at the tales and close to the mode of the pdf, and makes the discretisation around ± 1 (where f is changing the fastest) finer than required.

Finally, Fig. 3 shows the KL divergence from a standard Normal distribution to the discretised version found by Algorithm 1 (solid line) and [3] (dashed line). The number of intervals used during discretisation is given on the x-axis. For comparison, we also report the results found by simulated annealing (dotted line).

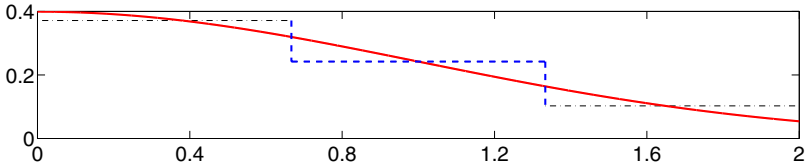
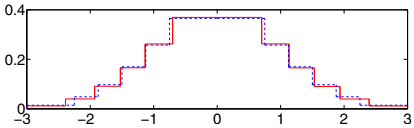
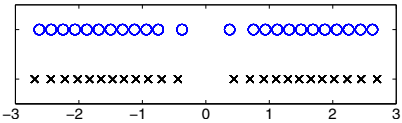


Fig. 1. Each interval is characterised by its discretised pdf



(a) Discretisation w/ 10 intervals



(b) Discontinuity points, 24 splits

Fig. 2. Part (a): The discretised pdf after (approximately) optimal discretisation (solid line) and the results of Algorithm 1 (dashed line). Part (b): The discontinuity-points chosen by Algorithm 1 (upper row, circles) and the approximately optimal solution found by simulated annealing (lower row, crosses).

3 Multivariate Distributions

Define the conditional KL divergence from one conditional distribution $f(\boldsymbol{y}|\boldsymbol{x})$ to another $\tilde{f}(\boldsymbol{y}|\boldsymbol{x})$ to be

$$D\left(f(\boldsymbol{y}|\boldsymbol{x}) \,\|\, \tilde{f}(\boldsymbol{y}|\boldsymbol{x})\right) = \int_{\boldsymbol{x}} f(\boldsymbol{x}) \int_{\boldsymbol{y}} f(\boldsymbol{y}|\boldsymbol{x}) \log \frac{f(\boldsymbol{y}|\boldsymbol{x})}{\tilde{f}(\boldsymbol{y}|\boldsymbol{x})}\, d\boldsymbol{y}\, d\boldsymbol{x},$$

so that we can calculate the KL distance between two joint distributions as

$$D\left(f(\boldsymbol{x},\boldsymbol{y}) \,\|\, \tilde{f}(\boldsymbol{x},\boldsymbol{y})\right) = D\left(f(\boldsymbol{x}) \,\|\, \tilde{f}(\boldsymbol{x})\right) + D\left(f(\boldsymbol{y}|\boldsymbol{x}) \,\|\, \tilde{f}(\boldsymbol{y}|\boldsymbol{x})\right).$$

Let us look at the behaviour of Algorithm 1 when we, for simplicity of exposition, consider the case where $\boldsymbol{\epsilon} = \emptyset$ (extending the results to the general case is straight forward), and define $\mathrm{pa}\,(z)$ to denote the parents of Z in the Bayesian network. Now it is easy to verify that the optimal discretisation $\bar{f}$ minimises

$$D\left(f(\boldsymbol{y}) \,\|\, \bar{f}(\boldsymbol{y})\right) = \sum_i D\left(f(y_i|\mathrm{pa}\,(y_i)) \,\|\, \bar{f}(y_i|\mathrm{pa}\,(y_i))\right),$$

Kozlov and Koller [3] obtains this by looking at the posterior joints at the clique level during their discretisation process. On the other hand, Algorithm 1 is looking at $\sum_i D\left(f(y_i) \,\|\, \bar{f}(y_i)\right)$, i.e., minimises KL-distances between *marginal distributions*, and thereby partly disregards the effect of the correlations between the random variables during discretisation, which can potentially result in an inferior discretisation. To investigate this further, we will now stress-test Algorithm 1 using a simple network consisting of only two nodes, $X \to Y$, but where the distributions are designed to make the models difficult for Algorithm 1.

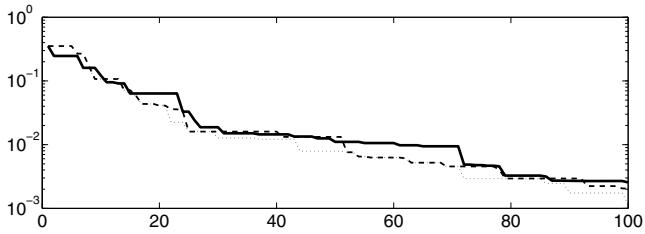


Fig. 3. The KL distance from the true pdf to the discretised pdf as a function of the number of intervals. The results of the "optimal" discretisation (found by simulated annealing) is given with a thin dotted line, and shown together with the results obtained using the true (dashed line) and approximated pdf (solid line) to find $f^{\downarrow}$and $f_{\ell}^{\uparrow}$ in the bound (Equation (2)). Note the log-scale on the y-axis.

3.1 Conditioning on a Uniformly Distributed Variable

Consider a Bayesian network as described above, let X follow a uniform distribution on $[0, 1]$ and let $Y|\{X = x\} \sim N(\mu = x, \sigma^2 = .1^2)$. Equation (2) will determine that there is no benefit from discretising X beyond the initial discretisation performed in Line 2 of Algorithm 1 (the pdf will remain at $\bar{f}(x) = 1$ independently of how many intervals the domain of X is discretised into). We can easily solve this by simply insisting on always refining the discretisation of every variable as the algorithm moves along, and use a heuristic like "*split the interval with the highest probability mass through its centre*" to guide our discretisation of X. The results in Fig. 4 have been obtained following this simple heuristic, and we present the conditional distribution of $X|\{Y = .1\}$. The correct result is a truncated Gaussian with its mode at 0.1, which is in good correspondence with the obtained results. The vanilla version of the algorithm fails to represent the correlation between X and Y.

3.2 The Resolution Problem

We now move to models that are "almost deterministic", meaning that the conditional variance $Y|\{X = x\}$ is small compared to the variance of X. We say that these models suffer from the *resolution problem*[3], and the difficulties occur as one tries to populate the conditional probability tables of the discretised model, i.e., while defining $P_{\mathcal{D}}(Y|X)$ in the discretised model $\mathcal{D}$.

Let us start by looking at why this problem arrises, before discussing how it can be solved. Consider a variable Z with mean μ_Z and standard deviation σ_Z, where we by Chebyshev's inequality have for any $\kappa > 0$ that

$$P\left(|Z - \mu_Z| \geq \kappa \cdot \sigma_Z\right) \leq \frac{1}{\kappa^2}. \tag{3}$$

[3] The problem was raised in the context of Algorithm 1 by Roger F. Sewell at Cambridge Consultants.

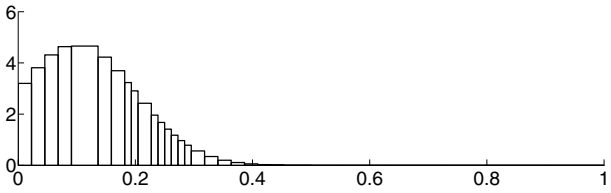


Fig. 4. Conditional distribution for $X|\{Y = .1\}$

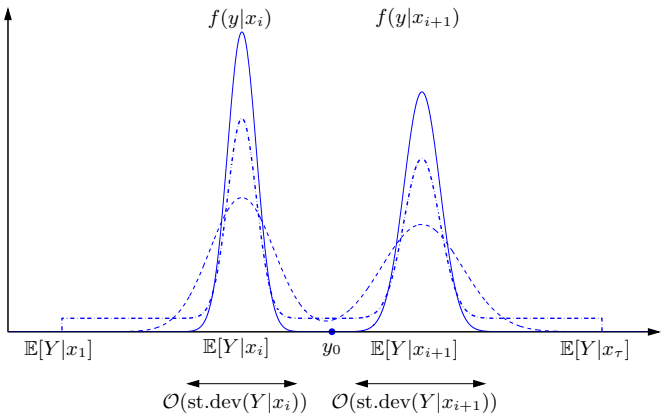


Fig. 5. The original conditional distributions $f(y|x)$ is evaluated for the two points x_i and x_{i+1}; shown with a solid line. Tempered versions of the conditional pdfs are shown with a dashed line, and the uniform fill-in approximation with dash-dots.

Assume now that we are looking to calculate the conditional probabilities that are required to define the discretised model. For the intervals $\omega_X = [\alpha, \beta]$ and ω_Y (defined for X and Y, respectively), we thus need to calculate $P(Y \in \omega_Y | x \in \omega_X)$. By simple manipulation we get

$$P(Y \in \omega_Y | x \in \omega_X) \propto \int_{y \in \omega_Y} \int_{x \in [\alpha,\beta]} f(y|x)\, dy\; f(x)\, dx,$$

which means that we for each $y_0 \in \omega_Y$ should calculate $\int_{x \in [\alpha,\beta]} f(y_0|x)\, f(x)\, dx$.

The integration will in general have to be done numerically, meaning that we will define a level of granularity, τ, an evaluation set containing τ ordered points $\{x_1, \ldots, x_\tau\}$ for which $\alpha = x_1 < x_2 < x_3 < \ldots < x_\tau = \beta$, and a set of constants (or "weights") $\{w_1, \ldots, w_\tau\}$. We then use the approximation

$$\int_{x \in [\alpha,\beta]} f(y_0|x)\, f(x)\, dx \approx \sum_{i=1}^{\tau} w_i \cdot f(y_0|x_i)\, f(x_i). \tag{4}$$

Here the values for w_i are chosen depending on which numerical integration scheme is employed (giving rise to composite versions of, e.g., the *rectangle rule*,

the *trapezoid rule*, or *Milne's rule* for numerical integration). Consider Fig. 5, where the conditional pds $f(y|x)$ are shown for $x = x_i$ and $x = x_{i+1}$ (solid line). The length of the area around $f(y|x_j)$ which contributes to the integral is of the order of the standard deviation of $Y|\{X = x_j\}$ (see Equation (3)). Notice that neither $f(y|x_i)$ nor $f(y|x_{i+1})$ contribute to the evaluation of the integral in Equation (4) for the choice of y_0 in this example. To evaluate Equation (4) with desired precision, a rule-of-thumb is therefore to choose τ such that

$$\tau \gg \frac{|\mathbb{E}[Y|X = \alpha] - \mathbb{E}[Y|X = \beta]|}{\min\{\text{st.dev}(Y|X = \alpha), \text{st.dev}(Y|X = \beta)\}}. \tag{5}$$

On the other hand, large values for τ can make the numerical integration extremely slow, and render the approach unsuitable for practical applications. This will be the case for models, which are affected by the resolution problem.

Our solution to performing the numerical integration in Equation (4) with sufficient precision using a limited calculation effort is inspired by *tempering* [6], which was developed to enable Markov chain Monte Carlo samplers to operate efficiently when confronted with multi-modal distributions. Let $\tilde{f}(y|\boldsymbol{x}, T)$ be the tempered version of the conditional pdf $f(y|\boldsymbol{x})$ at "temperature" T. Formally we define $\tilde{f}(y|\boldsymbol{x}, T) = \frac{1}{Z_T} \exp\left(\log\left[f(y|\boldsymbol{x})\right]/T\right)$, where Z_T is a function of T chosen to ensure that $\tilde{f}(y|\boldsymbol{x}, T)$ is a density. The tempered distribution is identical to $f(y|\boldsymbol{x})$ for $T = 1$, it approaches a uniform as T grows towards infinity, and can be seen as a smoothed version of $f(y|\boldsymbol{x})$ for T-values in-between. As an example we mention that the tempered version of a Gaussian with variance σ^2 at temperature T is a Gaussian with variance $T \cdot \sigma^2$. Fig. 5 includes tempered versions of $f(y|x_i)$ and $f(y|x_{i+1})$ drawn with a dashed line. As the tempered pdfs are "smeared out" versions of the true pdfs, they are easier to integrate (require a smaller granularity) than the originals. We note that it is not crucial to have an exact representation of the true pdfs during the first iterations of the discretisation procedure, when the goal is to roughly find which parts of a domain that contains probability mass. It is later, as the discretisation gets more refined, an accurate description of the true pdfs is required. However, at that time the problems with the numerical integration are also less prominent, because the discretisation has already reached an acceptable quality. Our plan is therefore to start the discretisation using a rather high temperature, and gradually cool the temperature towards unity as the discretisation gets more fine-grained.

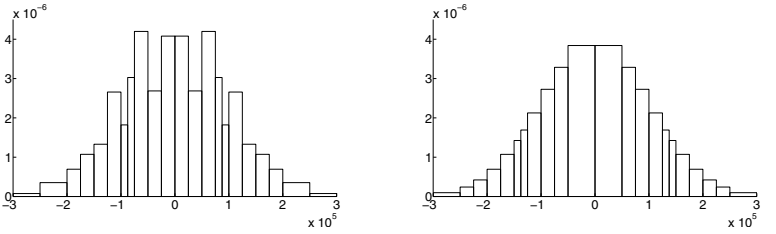
Unfortunately, tempering can be time-demanding in situations where the tempering distribution is not available in analytical form. We therefore approximate the tempering process by what we call a "*uniform fill-in*". The uniform fill-in of the conditional distribution $f(y|x)$ for $x \in [\alpha, \beta]$ is simply a mixture between the original distribution $f(y|x)$ and a uniform distribution on the interval $\big[\mathbb{E}[Y|X = \alpha], \mathbb{E}[Y|X = \beta]\big]$. The result of using a uniform fill-in is shown in Fig. 5 (dash-dotted line). It resembles tempering in that it "widens" the support of the integrand in Equation (4), thus initially allows the numerical integration procedure to operate with a lower granularity. As the discretisation is refined, less weight is put on the uniform, thereby enabling a faithful representation of

the true distribution as the discretisation process approaches its end. This is summarised in Algorithm 2, which takes the integration problem and the mixture probability as inputs. The filled-in distribution is defined as $\tilde{f}(y|x)$ in Line 2, where $U(y; a, b)$ is used to denote the pdf of the uniform distribution for Y on $[a, b]$. τ is then calculated according to the heuristic in Equation (5), but bounded above by a constant $\tau_{\max}$; note that it is $\tilde{Y}$ with pdf $\tilde{f}(y|x)$, which is used to find τ. The evaluation set and the weights are defined in Line 4; here giving a uniform spread and the trapezoidal rule, but other definitions are possible. Finally, the results are calculated in Line 5 using Equation (4).

1 **Function** `numericalIntegration`
  Input : Interval $[\alpha, \beta]$, pdfs $f(x)$ and $f(y|x)$, $y_0 \in \omega_Y$, mixture p.
  Output: Numerical approximation of $\int_{x\in[\alpha,\beta]} f(x)f(y_0|x)\,dx$.
2 $\tilde{f}(y|x) \leftarrow p \cdot U(y; \mathbb{E}[Y|X=\alpha], \mathbb{E}[Y|X=\beta]) + (1-p) \cdot f(y|x)$;
3 $\tau \leftarrow \max\left\{\tau_{\max}, \frac{\left|\mathbb{E}[\tilde{Y}|X=\alpha] - \mathbb{E}[\tilde{Y}|X=\beta]\right|}{\min\left\{\text{st.dev}(\tilde{Y}|X=\alpha), \text{st.dev}(\tilde{Y}|X=\beta)\right\}}\right\}$;
4 $\boldsymbol{x} \leftarrow \left[\alpha, \alpha + \frac{\beta-\alpha}{\tau-1}, \alpha + 2\,\frac{\beta-\alpha}{\tau-1}, \ldots, \beta\right]$, $\boldsymbol{w} \leftarrow \frac{\beta-\alpha}{\tau}\left[\frac{1}{2}, 1 \ldots, 1, \frac{1}{2}\right]$;
5 **return** $\sum_{i=1}^{\tau} w_i \cdot f(x_i) \cdot \tilde{f}(y_0|x_i)$;

Algorithm 2. Skeleton for the *uniform fill-in* algorithm

We will end this part by a small simulation study. Consider again the model $X \to Y$, and let X follow a Normal distribution with mean zero and variance σ_X^2. Further, let $Y|\{X = x\} \sim N(x, \sigma_Y^2)$. It follows that the marginal distribution for Y is $N(0, \sigma_X^2 + \sigma_Y^2)$. In Fig. 6 (a) we show the calculated marginal distribution of Y using Algorithm 1 with $\sigma_X^2 = 10^{10}$ and $\sigma_Y^2 = 10^{-6}$. We chose $\tau = 8$ in the numerical integration scheme, and conclude that the results are far from the sought solution. The results using adaptive selection of integration points (bounded above by $\tau_{\max} = 16$) and uniform fill-in are shown in Fig. 6 (b). The



(a)Vanilla version (b) Uniform fill-in and adaptive level selection

Fig. 6. The plots show the marginal distribution for Y after discretisation. Part (a) gives the results of Algorithm 1, whereas Part (b) shows the results when utilising uniform fill-in and the adaptive technique for choosing the level of the numerical integration

marginal density plot is quite close to the true model, and it is evident that the adaptions to the algorithm have removed the numerical problems the vanilla-version of the algorithm were suffering from.

4 Conclusions

In this paper we have discussed an approximative scheme for dynamic discretisation. The method was published in [5], and although it has been well received it has not been given a formal evaluation before now. Its distinguishing feature is that each variable is discretised based only on its marginal posterior distribution, which is in contrast to the original dynamic discretisation algorithm by [3], where the all variables in a clique are analysed together.

The simplification can potentially lead to remarkable speed-ups during inference (Kozlov and Koller's algorithm is about ten times slower than Algorithm 1 on an example network consisting of only two variables). On the other hand, we have explained why the results of the simplified approach may sometimes give inferior results. Sub-optimal discretisation is particularly evident for a variable having a large marginal variance compared to its conditional variance. We analysed this problem in detail and proposed a solution which significantly improves the quality of the results. In conclusion, our results suggest that the approximate procedure with simple extensions is feasible also for very challenging problems.

References

1. Shenoy, P.P., Shafer, G.: Axioms for probability and belief-function propagation. In: Proceedings of the Sixth Workshop on Uncertainty in Artificial Intelligence, pp. 169–198 (1990)
2. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Proceedings of the Twelfth International Conference on Machine Learning, pp. 194–202 (1995)
3. Kozlov, A.V., Koller, D.: Nonuniform dynamic discretization in hybrid networks. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, pp. 314–325 (1997)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, New York (1991)
5. Neil, M., Tailor, M., Marquez, D.: Inference in Bayesian networks using dynamic discretisation. Statistics and Computing 17(3), 219–233 (2007)
6. Neal, R.: Sampling from multimodal distributions using tempered transitions. Statistics and Computing 6, 353–366 (1996)

An Influence Diagram for the Collaboration in E-learning Environments

Antonio R. Anaya and Manuel Luque

Dept. of Artificial Intelligence, UNED, Juan del Rosal, 16. 28040 Madrid, Spain
{arodriguez,mluque}@dia.uned.es

Abstract. Influence diagrams have been used as decision support tool in different domains where the uncertainty plays an important role. The domain of collaborative learning environments can be characterized by the difficulties of proposing student collaboration indicators, and by the relationship between these indicators and the psychological and social student behavior. Thus, an analysis of the collaboration process muss take into account the natural uncertainty of the used indicators. For this reason we have built an influence diagram whose network has been created using the obtained findings in previous research. The influence diagram can support with a decision table that informs on the problematic circumstances of the target student to be recommended.

1 Introduction

Nowadays the educational institutions support students with e-learning environments [5], where the collaboration is possible and advisable [20]. Recommending to the students is a very hard task that requires frequent and regular analysis of the students' interactions [10,20]. When the collaboration has been assessed, it is necessary to determine: (1) who will be the person that will receive the recommendation, and (2) what will be the recommendation.

Influence diagrams (IDs) [9], as an extension of Bayesian networks, have demonstrated to be an effective tool for providing decision support for human experts in domains like medicine or engineering. They constitute a graphical representation that can help experts to structure the components of the domain, and to establish the quantitatively the relationship among the components using probability. These technologies are used in context to deal in the uncertainty. We noted that collaborative learning depend on the social and psychological student behavior [10], which is out of control in a typical e-learning environment. Thus, the collaboration analysis must deal in uncertainty and probability.

For these reasons we have created na ID to ask the previous questions. The ID's nodes are the attributes that tracked the students' interactions and assessed the students' collaboration, and the arcs have been chosen according to the semantics of the attributes selected. The probabilities were automatically learned from the dataset recording the students tracking and assessments, and the classifications by the collaborative learning experience expert.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 235–244, 2013.

When the ID was built, an evaluation algorithm was used to compute an optimal policy for the decision variable. The policy took the form of a table that contains an exhaustive table of the state space of the variables observed when making the decision. Due to this table, an automatic recommender can send a warning to target student or teacher to inform on the detection of a collaboration problem.

In the next section we explain the previous works, on which this research is based, that is, the literature on the indicators of collaborative learning and influence diagrams. In Section 3 we describe the construction of the ID, the used variable of the students collaboration, the structure of the network and the elicitation of the probabilities and utilities. Following we show the results of the application of the proposed ID with a real data from collaborative learning experience. The paper finishes with the conclusions and futures works.

2 Previous Works

2.1 Collaborative Learning Indicators

To be able to assess the collaborative process to know whether the collaboration learning takes place, it is necessary a regular and frequent analysis and tracking of the students' interactions in the collaborative learning environment. Collaboration is based on communication and thereby to improve collaborative learning, collaboration analysis must analyze communication. Forums are key communication tools which are widely used in e-learning environments to support student collaboration [11]. [6] analyzed forums and revealed student characteristics quantitatively without encoding the messages from senders. These authors also researched the effect of time variables in the analysis of forum interactions to support the collaborative learning process [7].

Some researchers have analyzed interactions in forums to improve collaborative learning in terms of student characteristics that were related to collaboration. For instance, [21] and [8] proposed indicators from interactions in forums as the number of messages or the number of conversations started. They proposed an explicit relationship between these indicators and student activity and initiative. Moreover, [18] proposed expert analysis of student messages in forums to identify topics and labeled these topics. Experts' involvement in the analysis process hampers the possibility of getting frequent and regular analysis of students' interactions, which may help them in further their collaboration.

As for the analysis method, different approaches were proposed to analyze student collaboration in e-learning environments. We focused on data mining (DM) techniques, which have been used to discover knowledge that is hidden in student interactions [19]. Our research focus has been to find quantitative collaboration analysis with transferability features which supports timely assessment of student collaboration in current e-learning environments. As it is confirmed from the state of the art analysis, forums are key tools in supporting collaboration and can be analyzed to reveal relevant student collaboration characteristics. To this DM approaches can support the analysis with minimal teacher or student

intervention. Bearing this in mind in our previous research, we proposed two different approaches to assess student collaboration [1]. Both approaches used indicators of the student interactions in forums as the data source.

2.2 Bayesian Networks and Influence Diagrams

Bayesian networks (BNs) [16] constitute a modeling framework where probabilistic dependence and independence relations are easily represented. One important feature of Bayesian networks is that they allow an easy communication with the expert when the model is being built and debugged [12].

Influence diagrams [9] are an extension of Bayesian networks to include decision variables and the preferences of the decision maker. Influence diagrams allow the expert not only to perform diagnosis tasks as in Bayesian networks, but they also provide with very interesting possibilities [4]: (1) to know which is the best action in each scenario, (2) obtain automatically explanations of the reasoning of the system, and (3) formulate queries about the posterior probabilities and expected utilities of some variables of the model.

Formally, an influence diagram (ID) consists of an acyclic directed graph $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the *set of nodes* and $\mathbf{E}$ is a set of directed edges between nodes in $\mathbf{V}$. The set $\mathbf{V}$ is partitioned into three disjoint sets: chance nodes $\mathbf{V}_C$ (circles or ovals), decision nodes $\mathbf{V}_D$ (graphically represented by squares or rectangles), and utility nodes $\mathbf{V}_U$ (diamonds or hexagons). Chance nodes represent uncertain events. Decision nodes represent the actions under the direct control of the decision maker. Utility nodes represent the preferences of the decision maker. Each node corresponds to a variable.

There are three types of arcs in an ID, corresponding to the type of node they go into. Arcs into chance nodes represent probabilistic dependency. Arcs into decision nodes, called *informational arcs*, represent availability of information; i.e., if there is an arc from a node X to a decision node D then the state of X is known when decision D is made. Arcs into utility nodes indicate the domain of the associated utility function.

We assume all chance and decision variables of the ID are discrete and have finite domain. Given a chance or decision variable X, we define the *domain of* X, denoted $dom(X)$, as the set of possible values of X. The *state space* of a set of chance or decision variables $\mathbf{X}$, denoted $space(\mathbf{X})$, is defined as the cartesian product of the domains of the variables in $\mathbf{X}$. A *configuration* is an element of $space(\mathbf{X})$. The set of variables known when making a decision D is the set of *informational predecessors* of D, and denoted by $iPred(D)$.

The quantitative information that defines an ID is given by (1) assigning to each chance node C a conditional probability potential[1] $P(c|pa(C))$ for each configuration of its parents, $pa(C)$[2], and assigning to each utility node U a potential $\psi_U(pa(U))$ that maps each configuration of its parents onto a real number.

[1] A *potential* is a real-valued function over a finite domain of variables.

[2] We denote with $Pa(X)$ the set of parents of X, and with $pa(X)$ a configuration of the parents of X.

A *policy* for a decision D in an influence diagram is a function that maps each configuration of the variables in $iPred(D)$ onto a value of D. The main objective of the evaluation of an influence diagram consists in finding an *optimal strategy*, which maximizes the expected utility and contains an optimal policy for each decision.

3 System Description

3.1 Variables

We have noted that we focus on the students interactions in the forums of a collaborative learning experience to assess frequently the collaboration learning. But we should describe briefly the collaborative learning experience so that the audience understand the choice reason of the variable that we use in the ID.

A collaborative learning experience has been offered to students during five academic years (2006-07, 2007-08, 2008-09, 2010-11, 2011-12) and it can be divided into two phases [2,1]. In the first phase the students have to make an individual task, and a learning platform is offered students so that they can work and communicate each other. In the second phase the students are grouped into three members teams. In this phase a private space in the learning platform is offered every team. In this private space enough services are supported to work and collaborate (forums, news, task manager, FAQ's, etc.). The team members have to make five collaborative tasks and can use the forums of the private space to manage the team work. We have used a set of the indicators of the interactions in forums of the teams' private space. We have proposed indicators of the students interactions in forums because these indicators can be obtained frequently without a human intervention but they are capable of describe students characteristics and behavior [6]. The indicators were explained in depth in [2].

In this research we proposed a set of indicators, which are based on those from the previous research, to track the student's initiative, activity and regularity, and acknowledgment by their peers. Some indicators are related to the student's initiative (N_{thrd}) and the regularity of the initiative (L_{thrd}) , others to the activity (N_{msg}) and the regularity of the activity (L_{msg}), and the third group of indicators are related to the acknowledgment (N_{r_thrd}, N_{r_msg}) of the students by the peers. We calculated the indicators N_{thrd} and N_{msg}, which are connected to a period of time (one day). For this reason, the variance is related to the regularity. If all students had the same value of N_{thrd} or N_{msg}, lower values of variance would indicate a higher regularity. In the case of the indicators measuring the replies to student interactions, acknowledgment can be measured by the replies to student initiative N_{r_thrd} or activity N_{r_msg}. These indicators were chosen due to their understandability and the fact that they are relevant for the problem that we are considering.

In addition of the quantitative tracking, an expert's analysis was done. An expert, who was the tutor of the collaborative learning experiences, read the forum messages of all the students who participated in the collaborative learning

experiences, and labeled students according to their collaboration into three labels (high, medium, low).

We also proposed two DM approaches, which used the indicators of the students interactions, to assess the students' collaboration. The different approaches were: *the Clustering Approach* and *Metric Approach*. Both approaches are explained in depth by [2]. The *Clustering Approach* consisted of grouping of students indicators into three cluster by EM clustering algorithm. We proved previously that the clusters indicates the students collaboration level assessment (*Level*). The *Metric Approach* is basically a metric, mathematical formula, was built using the indicatorsL_{thrd}, L_{msg} and N_{r_msg} and offered a collaboration assessment *Metric*. The metric's indicators were related strongly to the students collaboration by means of a machine learning process based on decision tree algorithms.

Finally, the variables corresponding to the selected indicators are: related to the students' initiative *(*N_{thrd}*)*; about the regularity of the initiative (L_{thrd}); related to student's activity (N_{msg}); regularity of the activity *(*L_{msg}*)*; students' acknowledgment *(*N_{r_thrd} and N_{r_msg}*)*. Other two variables that give information about the global collaboration, which were inferred by the DM approaches, are: collaboration level *(Level)*; and collaboration Metric *(Metric)*. These variables are chance variables, because they represent uncertain events that are not under the direct control of the decision maker. The expert-based analysis collaboration label is represented by variable *Collaboration*. This variable is a hidden attribute, whose certain value is unknown, but it can be indirectly observed by means of the student's interactions. The decision of the problem is represented by variable D. It indicates if an intervention is necessary or not to solve the collaboration issues detected by the system. The preferences of the decision maker are represented by the utility variable U.

3.2 Domain of the Variables

We had to determine the possible values for each variable. Chance variables of our problem can be interpreted as continuous quantities, but we decided to define them as discrete mainly because the software tool used for the construction and the inference only admitted discrete variables.

We had to decide the level of *granularity* when discretizing the chance variables. For the discretization we have tried to find an equilibrium among (1) the precision of the model, (2) the difficulty of the parameters's elicitation and the interpretation of the results, and (3) the understandability of the values by the someone not expert. All chance variables are ternary in a scale, with values *high*, *medium* and *low*.

The indicator *Level* is discrete and its possible values are: *high*, *medium* or *low*. The other indicators have been discretized by splitting the numerical scale of the indicator possible values among the lowest and the highest into three intervals. We assigned the discrete value *low*, *medium* or *high* to the subsequent interval.

Decision variable D is discrete, with two possible states (*yes* and *no*) corresponding to the actions that can be chosen by the decision maker: *yes* indicates that an action must be taken to solve the collaboration problem, while *no* indicates no action is required. Utility variable U is continuous, and it is assumed to take values ranging from 0 to 100, which is a scale proposed by the tutor of the collaborative learning experience.

3.3 Structure of the Graph of the Influence Diagram

Each variable identified in the problem has been represented with a node in the graph of the influence diagram (see Figure 1). The structure of the graph have been built by following the expert's indications.

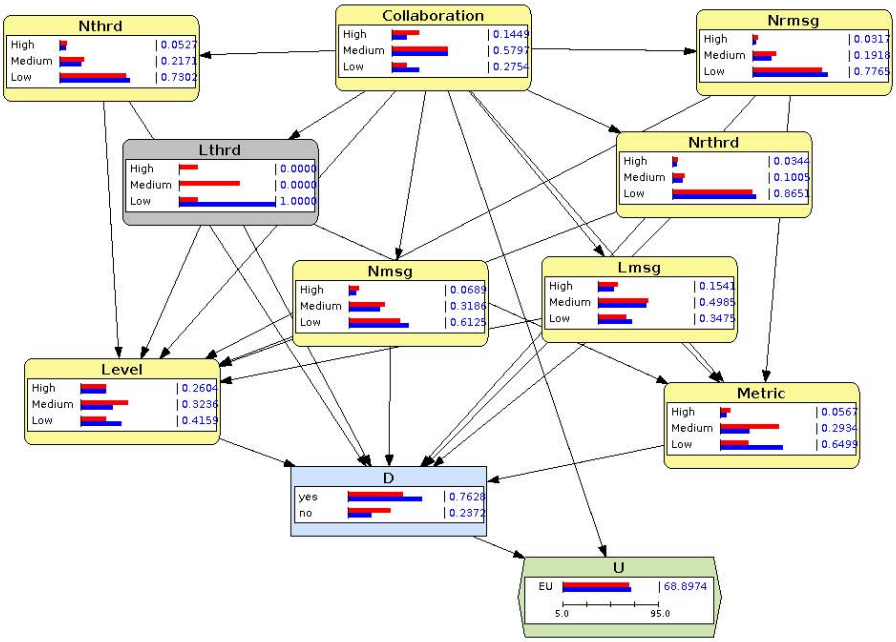


Fig. 1. Influence diagram model for the collaboration in e-learning environments

Node *Collaboration* is the target variable in the graph [14]. Attributes representing statistical indicators (N_{thrd}, L_{thrd}, N_{msg}, L_{msg}, N_{r_thrd} and N_{r_msg}) and assessments indicators (*Level* and *Metric*) have a causal and probabilistic relationship with *Collaboration*. That justifies the arcs pointing from *Collaboration* to each of these attributes. Moreover, while the indicators' attributes are probabilistically independent given the value of variable *Collaboration* [17], attributes *Level* and *Metric* also depend on other indicators. That is also reflected by drawing the corresponding arcs.

Attribute D is the only decision in the diagram. Given that variable *Collaboration* is not observable, it is unknown for the decision D and there is therefore no arc from the node *Collaboration* to node D. However, variables N_{thrd}, L_{thrd}, N_{msg}, L_{msg}, Nr_{thrd}, N_{r_msg}, $Level$ and $Metric$ are observable, their values are known when making decision D and can therefore be observed by the decision maker in that moment. That justifies the arcs pointing from those nodes to the node D.

The parents of U in the diagram, which are those defining the domain of its utility function, are *Collaboration* and D.

3.4 Elicitation of Numerical Parameters

In order to have the ID model it is necessary to specify the numerical parameters of probability potentials and the utility function. For example, for each chance node C we must give a conditional probability potential $p(C|pa(C))$ for each configuration of its parents, $pa(C)$. However, but given the restriction that $\sum_c P(c|pa(C)) = 1$, only some of the parameters are independent.

With regards to the specification of probability parameters, every chance variable has required the elicitation of 6 probability parameters, except the next variables: *Collaboration* (2), *Metric* (54), and *Level* (4374). It is remarkable than while the number of parameters elicited in the ID is high (4436), most of them correspond to the probability table of variable *Level*.

The numerical parameters of probability tables have been automatically learned from a data base cases file. The data base cases file did not contain all the necessary registries for the computation of the entire set of parameters of conditional probabilities, and that makes necessary to use Laplace's correction [15] to alleviate the problem.

The utility function has required the specification of 6 parameters. To elicitate the utilities, we have identified three clear scenes in our problem: when a student has a high, medium or low collaboration. Therefore, the recommendations should increase the student's collaboration in each scene or, at least, should not decrease the student's collaboration. To elicitate the utilities we took into account the three scenes of the student's collaboration and the advice given by the tutor of the collaborative learning experience. The scale of the utilities, from 0 to 100, has been chosen because it is understandable by the tutor of the collaborative learning experience.

- In the case the student's collaboration is high, the recommendations should not bother the student, because s/he is doing good work, but encourage sometimes. For this reason we have proposed utility elicitations in this case as: 90 not recommend and 10 recommend.
- In the case the student's collaboration is medium, the recommendation should inform student that the circumstances are enough good but these circumstances could get worse. Because the student works in team correctly, the recommendation can not bother him/her, of course. Thus, we have proposed 40 not recommend and 60 recommend.

– In the case the student's collaboration is low, the recommendation should promote the student's interactions are increased. We have proposed 5 not recommend and 95 recommend. We have not proposed to recommend always because the possible uncertainty.

The elicitations of the utilities can be checked comparing the attributes of the student's interactions and collaboration assessments before and after the recommendation. Therefore, we can consider that one recommendation is useful, in general, when the interactions are increased.

4 Results

After the phase of construction of the system we have an ID available in the software OpenMarkov [3]. The ID has been evaluated to obtain the maximum expected utility and the optimal policy of the decision D. Figure 2 shows a small part of the policy table of decision D obtained with OpenMarkov. The policy table is an important element of the explanation of the reasoning of the system [13]. It offers to the expert the possibility of examining every configuration (sets of observations) known by the decision D and to see what is the best action recommended by the system (cells red colored in the table with the number 1.0).

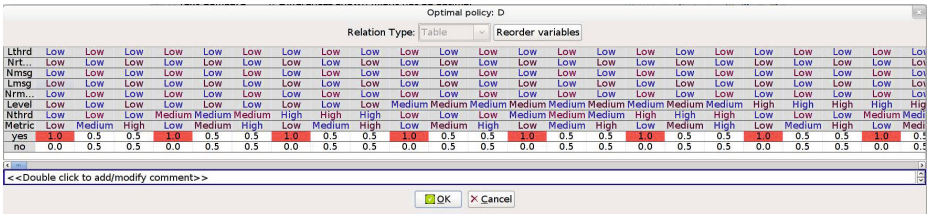


Fig. 2. Optimal policy of decision D

Moreover, the ID integrates all the information of the model (variables, relations and numerical parameters) in a compact representation that eases enormously any possible modification. Moreover, the debugging phase of the model can benefit from the explanation capabilities of the software tool available for IDs. In our case, the ID in OpenMarkov offered us the possibility of analyzing different scenarios of observations and to see how the posterior probabilities of the target variable *Collaboration* are modified according to the evidence introduced in the GUI. For example, in Figure 1 the expert is analyzing two *evidence cases*: (1) the situation in which we do not have evidence on any variable, and (2) the situation in which we observe that the value of variable L_{thrd} is *low*. The expert can see how the probabilities are different in both cases, and to see that the probability of $D = yes$ is higher in the second case due to the observation of L_{thrd}.

5 Conclusions and Future Works

In this paper we have presented an ID that works as decision support tool in a collaborative learning context. The ID's network has been created using as attributes or nodes of the network the indicators of the students interactions in forums of a collaborative learning experience. The tutor helped to structure the network and the data from previous experiences were used to elicitate the probabilities of the network.

Finally the ID have offered a decision table that informs on the appropriate circumstances, in oder works, the configuration of the attributes values, which represent the students interactions in forums, to recommend. The ID works without the intervention of the tutor of the experience and while the collaborative learning experience is running. The ID informs tutor about the students and their circumstances to be recommended.

The first future works is to build an recommender that warms automatically tutor and target student on the problematic situation in the collaboration process. In order to gain the acceptance by the expert, the tutor, two tasks can be performed with the decision table. First, the analysis the decision table to discover the most relevant attributes in the decision, which can inform tutor on the most relevant attributes to identify the problematic circumstances. Second, a graphical representation of the policy can be built to represent visually the students' circumstances of recommendation, which will allow the expert to understand better the recommendation problem and provide her a general vision of the recommendation scenes.

References

1. Anaya, A.R., Boticario, J.G.: Application of machine learning techniques to analyze student interactions and improve the collaboration process. Expert Systems with Applications 38(2), 1171–1181 (2011)
2. Anaya, A.R., Boticario, J.G.: Content-free collaborative learning modeling using data mining. User Modeling and User-adapted Interaction 21(1-2), 181–216 (2011)
3. Arias, M., Díez, F.J., Palacios, M.P.: OpenMarkovXML. A format for encoding probabilistic graphical models. Technical Report CISIAD-10-04, UNED, Madrid, Spain (2010)
4. Bielza, C., Gómez, M., Shenoy, P.P.: A review of representation issues and modelling challenges with influence diagrams. Omega 39, 227–241 (2011)
5. Daradoumis, T., Juan, A.A., Lera-López, F., Faulin, J.: Using collaboration strategies to support the monitoring of online collaborative learning activity. In: Lytras, M.D., Ordonez De Pablos, P., Avison, D., Sipior, J., Jin, Q., Leal, W., Uden, L., Thomas, M., Cervai, S., Horner, D. (eds.) TECH-EDUCATION 2010. CCIS, vol. 73, pp. 271–277. Springer, Heidelberg (2010)
6. Dringus, L.P., Ellis, E.: Using data mining as a strategy for assessing asynchronous discussion forums. Computers & Education 45, 140–160 (2005)
7. Dringus, L.P., Ellis, E.: Temporal transitions in participation flow in an asynchronous discussion forum. Computers & Education 54(2), 340–349 (2010)

8. Gaudioso, E., Montero, M., Talavera, L., Hernandez del Olmo, F.: Supporting teachers in collaborative student modeling: A framework and an implementation. Expert Systems with Applications 36, 2260–2265 (2009)
9. Howard, R.A., Matheson, J.E.: Influence diagrams. In: Howard, R.A., Matheson, J.E. (eds.) Readings on the Principles and Applications of Decision Analysis, pp. 719–762. Strategic Decisions Group, Menlo Park (1984)
10. Johnson, D.W., Johnson, R.: Cooperation and the use of technology. In: Handbook of Research on Educational Communications and Technology, pp. 401–424. Taylor & Francis, Abington (2004)
11. Jordan, L.E.: Transforming the student experience at a distance: Designing for collaborative online learning. Engineering Education: Journal of the Higher Education Academy Engineering Subject Centre 4(2) (2009)
12. Lacave, C., Díez, F.J.: A review of explanation methods for Bayesian networks. Knowledge Engineering Review 17, 107–127 (2002)
13. Lacave, C., Luque, M., Díez, F.J.: Explanation of Bayesian networks and influence diagrams in Elvira. IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics 37, 952–965 (2007)
14. Millán, E., Loboda, T., Pérez de-la Cruz, J.L.: Bayesian networks for student model engineering. Computers & Education 55(4), 1663–1683 (2010)
15. Niblett, T.: Constructing decision trees in noisy domains. In: EWSL 1987, pp. 67–78 (1987)
16. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988)
17. Pearl, J., Geiger, D., Verma, T.: Conditional independence and its representations. Kybernetika 25, 33–44 (1989)
18. Perera, D., Kay, J., Yacef, K., Koprinska, I.: Mining learners' traces from an online collaboration tool. In: Proceedings of the 13th International Conference of Artificial Intelligence in Education, Workshop Educational Data Mining, Marina del Rey, CA. USA, pp. 60–69 (July 2007)
19. Romero, C., Ventura, S.: Educational data mining: A review of the state-of-the-art. IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews 40(6), 601–618 (2010)
20. Swan, K., Shen, J., Hiltz, S.R.: Assessment and collaboration in online learning. Journal of Asynchronous Learning Networks 10(1), 45–62 (2006)
21. Talavera, L., Gaudioso, E.: Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In: Proceedings of the Workshop on Artificial Intelligence in CSCL, 16th European Conference on Artificial Intelligence (ECAI 2004), Valencia, Spain, pp. 17–23 (2004)

A Bayesian Model for Lexical Availability of Chilean High School Students in Mathematics

Pedro Salcedo L.[1], Anita Ferreira C.[2], and Fernanda Barrientos C.[2]

[1] Research and Educational Informatics Department
[2] Department of Spanish Linguistics
University of Concepción, Chile
{psalcedo,aferreir}@udec.cl

Abstract. Lexical availability studies make an account from a statistical point of view about the lexicon of a group of people and the extent to which these might be considered representative of a bigger population. Recent findings in lexical availability prove the existence of patterns in order and word choice when different individuals recall words after a certain category or center of interest is given as a stimulus, and that these patterns can be modeled as a graph. The following article goes beyond the common idea of lexical availability in terms of frequency, stating that lexical availability can be modeled from a probabilistic perspective, i.e. calculating the probability that a certain word is available given other previous words. This work also has implications in prototype semantics theory, as it provides a mathematical model to obtain structures of the mental lexicon that can be extrapolated to the members of specific groups of people.

Keywords: Lexical availability, Bayesian networks.

1 Introduction

Although studies in lexical availability began in the 50's and were aimed to determine a basic vocabulary of French as a second language, nowadays the concept has been mainly researched and its scope and applications have varied significantly. An important application is in the field of education, from which many hypotheses have been posited in terms of its influence in learning different subjects and not only a second language. It has been stated that students with more available words in a certain field are more successful in their studies [7].

This work is part of a bigger research project (Fondecyt 1120911) [19] whose main goals are a) to quantify and describe the mental lexicon of high school students in the area of Mathematics regarding the following centers of interest: numbers (arithmetic), algebra, geometry, probabilities and chance (statistics) and other more common centers of interest (transportation, home and city); and b) to design a pedagogical proposal to generate an Adaptive Hypermedia System that allows for an increase in lexical availability in the centers of interest named above.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 245–253, 2013.

2 Theoretical Framework

2.1 Studies in Lexical Availability

Mental Lexicon. The concept of mental lexicon refers to a mental state of the knowledge about words [6], a concept that attempts to explain the way that words are stored inside the human mind. Initially, the lexicon was understood as a static storage of different lexical items; however, as interest in the concept of mental lexicon increased in the fields of psycholinguistics and cognitive science, a more dynamic model emerged: today it is widely accepted that the mental lexicon is a complex group of networks consisting of lexical items that are interconnected and subject to change, as new words are learned, others are forgotten, and others are replaced [6]. In this respect, studies in lexical availability aim to measure the ease with which a word can be recalled by a specific person when asked about a certain category [8], which allows a better understanding of how words are connected to each other. However, more recent views of the mental lexicon do not only consider the connections that can be made among words from a certain category, but also the connections that can be made among words according to world knowledge [5].

Lexical Availability. Lexical availability studies began in the 50's, given the need to determine a basic vocabulary to teach French as a second language. Michéa (1953; cited in [4]) coined the term and defined it as a word that might not be particularly frequent, but is stored in the speaker's lexicon and can be recalled immediately when needed. Hence, the availability of a word can be measured as soon as a stimulus is presented to a certain individual. Usually, studies in lexical availability present a category (or centers of interest) to individuals who are members of a certain population and are then asked to mention or write down all the possible words that they can recall that are related to that category. The lists will show not only the words that individuals are able to associate to a certain category, but also the order in which they are recalled; hence, the availability of a certain word that showed up in first place in a list is not the same than another word in fifth place.

The resulting word lists are then analyzed with the formula proposed by López Morales (1978; cited in [4]) that was slightly adjusted to a more intuitive analysis by Echeverría et al [4]. The final formula is as follows:

$$D(p) = \frac{f_1 + \lambda f_2 + \lambda^2 f_3 + ... + \lambda^{n-1} f_n}{N}$$

However, it is worth mentioning that even though it is same person who performs this task, results can be different [6]. This is considered as a proof that the lexicon is mostly dynamic, and therefore the words are not stored in fixed places with a fixed recall time, or importance, or order of appearance.

Later, Echeverría et al. [3] considered that although these word lists obtained differed from each other, it was possible to find a pattern of lexical order. This

lead them to propose a lexicon model with the data obtained from their previous studies in lexical availability, in which the items of the different word lists were analyzed through a computational application. The program would form graphs in which the nodes represented the words and the edges had values that represented the number of times that a specific word was followed by another certain word. For instance, if the word uva (grape) was followed by the word durazno (peach) in two word lists, then the program would generate a graph with two nodes linked by an edge with a value of 2.

Our aim is to build on the model presented by Echeverría by adding a probabilistic component, thus being able to more accurately predict the mental structures in a particular group of people.

2.2 Bayesian Networks

When modeling environments or phenomena in which no direct causality relationships can be found, probability-based models can predict with a fair amount of precision the possible outcomes of an event. The Bayes Theorem calculates the probability that a certain event occurs when other previous events have occurred, by which it is possible to make an inference. A Bayesian network is a graphical model that finds probabilistic relationships among variables of a system [1] by using the Bayes formula.

According to graphs theory, a Bayesian network is a acyclic directed graph in which the nodes represent different variables and the edges represent the relationship among them, whether they are informational or causal [14]. The following is the simplest Bayesian network, in which A represents the parent node, and B is a child node. The edge pointing to B is interpreted as A being the previous event and B the second event, implying a causal relationship (A causes B).

Fig. 1. A simple Bayesian Network with one parent node (A) and one child node (B)

Therefore, a Bayesian network represents either the joint probability of a certain event to happen, given one or more previous events; or the probability of various events to happen, given one single previous event. In order to build a Bayesian network, some parameters are required: first, the prior probability of the event represented by the parent node; and secondly, the probability of A

given event B. After this, the probability of the presence of event A given B (i.e. $P(\frac{+a}{+b})$ can be calculated through the following formula:

$$P(X_1, .., X_n) = \prod_{i=1}^{n} P(\frac{X_i}{pa(X_i)})$$

Linguistics has been using Bayesian networks to model different aspects of human language, such as comprehension and production in general [12] or more specific aspects of it such as human sentence processing [11], and speech processing [10]. We attempt to understand semantic relationships among words as a probabilistic model, in which the order of available words is the subject of probability, as they appear only when others have been previously recalled.

2.3 Lexical Availability in Mathematics

Currently one of the most significant problems in mathematical education comes from the difficulty that students have in reading the wording in the discipline, which leads naturally to relate language to the process of learning of this science [17].

The act of communicating in "Mathematics" is related to a particular language, where every word must correspond to a communicative act to avoid errors in the construction of mathematical meanings ([13]; [20]; [15]; [2]; [16]).

It follows from the importance of knowing the mathematical lexicon latent in a particular group and model this lexicon so as to predict the probability that a word or construct appears according to the lexicon already acquired ([17]; [18]).

3 Bayesian Modeling of Lexical Availability

Echeverrías work went towards a model for the mental lexicon that could be representative of a certain group of people, supported by previous studies in lexical availability and graphs theory. The model that we propose takes Echeverras research as a framework, in order to model lexical structures that can be considered as representative of a certain group of people; in this case, the data corresponds to Chilean high school students and the center of interest is only related to Mathematics. However, our point of view includes a probabilistic component, which considers the existing lexicon in a certain group as well as its social background, which are described in the following section.

Our model calculates the probability that a certain word is available in a given populations lexicon given other variables: previous word(s) elicited, type of school where the student is enrolled, and sex of the student. As we already stated, the probability of a certain event when other previous events have also happened is given by the product of the product of such event given the probabilities of the parent nodes. Notice that in order to calculate the probability of a certain word that has been elicited after other word(s), then an iteration process is needed.

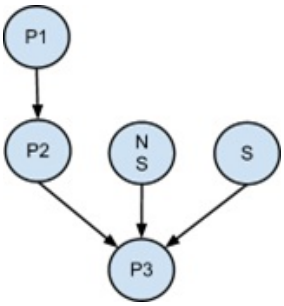


Fig. 2. A Bayesian model to predict the presence of a certain word in the lexicon of a specific population, given by P(p3 — pa(p3)) = P(p2 — p3) P(p3) P(p2) x P(NSa—p3) P(p3)P(NSa), etc

4 Methodology

To test the above model we used the data collected in the project FONDECYT 1120911 [17], for which we used the following methodology.

A test was applied in order to measure the latent lexicon, considering the student's social profile (sex, age and social background) which subsequently determine the frequencies of occurrence of each Word (node in the graph) and the frequency of relationships between each of them, allowing later analysis using the proposed model.

The test: This consisted of a form divided into two parts. The first was an identification item, in which subjects had to write down their name, age, sex and year of studies. In the second part of the test, subjects had two minutes for each center of interest to write down all the words that they could recall. .

Hypothesis: The proposed Bayesian model is effective at predicting the onset of lexical units in the students' semantic network, in the field of mathematics. .

Sample: It consisted of 1500 subjects, from three types of educational institutions in Concepcin, Chile (Municipal, fully funded by the government; Subvencionado, funded only in part by the government; and Particular, private), each of which has a specific social backgrounds. .

Variables: Since this study aims to test the proposed Bayesian model, it is necessary to consider the richness of the lexicon of every social background, thus considering as variables, sex, social class, and the number of words elicited by center of interest.

Sample edition:. In order to validate the data, the following decisions were taken: type all in caps; using singular form in nouns and adjectives; refer diminutives to the primitive word; and refer all verb forms to infinitive, except gerund and participle. Compound words were entered together by using hyphens.

To determine the words that belong to each center of interest, experts determined the words related to the progress maps in each area. Thus, all those words unrelated to mathematics and the specific center of interest were filtered.

Lastly, this study took only the first 50 words of the lexicon belonging to every social background and each specific center of interest with the largest number of connections (more than 5) and the highest frequency (greater than 8). This allowed nodes without significant influence on others to be left.

Data Processing: The data was computationally processed to calculate frequencies and probability. To determine the influences ELVIRA system was used (Fig. 3)

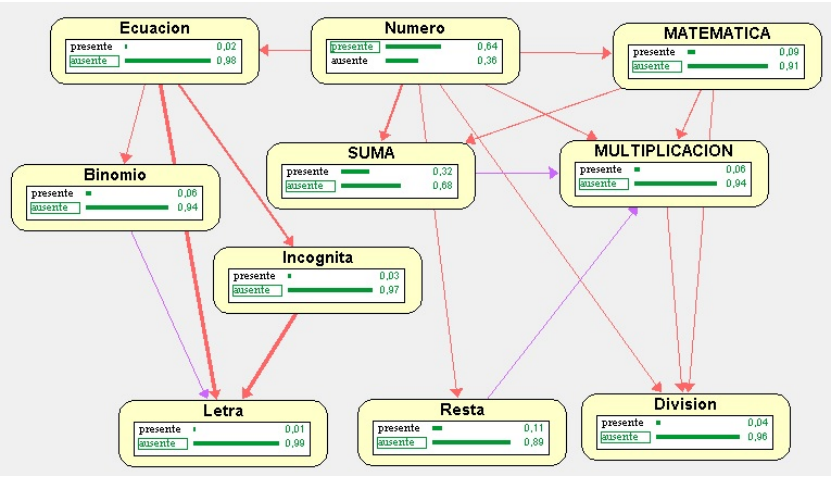


Fig. 3. Network developed with ELVIRA

5 Results

In order to test the model, only 10% of the individuals in the sample (150 students) was chosen, ignoring irrelevant nodes and edges (with a frequency lower than 8). The percentage that the probability of appearance of each node (word) shown by the model matches the words presented by the subject has been determined.

Table 1 shows an example of the first 10 words that were most mentioned by the students and the corresponding percentage, in the specific center of Algebra and considering social backgrounds from working class to upper class (Municipal, Subvencionado and Particular).

The frequency of the relationships among words is determined by the sequence in which students mentioned each word in the survey. Table 2 shows the number of times a word is mentioned after another.

Finally, to test the model percentage of certainty, a 10% of the whole sample has been used, which is compared with the prediction results of each of the words, obtaining table 3, which shows that on average between the three types of educational institutions, the model has a 65.

Table 1. The 10 most frequently mentioned words in the survey, with the corresponding relative frequencies

	MUNICIPAL	Fr Abs	Fr %	SUBENCIONADO	Fr Abs	Fr %	PARTICULAR	Fr Abs	Fr %
1	NUMERO	55	11,2%	NUMERO	159	6,6%	NUMERO	47	7,2%
2	LETRA	30	6,1%	LETRA	129	5,4%	LETRA	42	6,4%
3	MATEMATICA	24	4,9%	SUMA	114	4,7%	ECUACION	24	3,7%
4	SUMA	22	4,5%	RESTA	105	4,4%	MULTIPLICACION	23	3,5%
5	MULTIPLICACION	20	4,1%	MULTIPLICACION	99	4,1%	DIVISION	22	3,4%
6	DIVISION	18	3,7%	DIVISION	81	3,4%	RESTA	20	3,1%
7	RESTA	18	3,7%	MATEMATICA	73	3,0%	SUMA	19	2,9%
8	SIGNO	17	3,5%	ECUACION	68	2,8%	X	18	2,8%
9	PROBLEMA	13	2,7%	FRACCION	61	2,5%	INCOGNITA	17	2,6%
10	FRACCION	12	2,4%	BINOMIO	55	2,3%	FRACCION	13	2,0%

Table 2. Percentage of appearance of a word in a sequence

	DIVISION	FRACCION	LETRA	MATEMATICA	MULTIPLICACION	NUMERO	PROBLEMA	RESTA	SIGNO	SUMA
DIVISION	0,0%	0,0%	1,4%	0,0%	0,0%	0,0%	0,0%	0,7%	0,7%	0,7%
FRACCION	1,4%	0,0%	0,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
LETRA	0,0%	0,0%	0,0%	0,7%	0,0%	0,0%	0,0%	1,4%	0,7%	6,1%
MATEMATICA	1,4%	0,0%	0,0%	0,0%	2,0%	2,7%	0,0%	0,0%	0,0%	0,0%
MULTIPLICACIO	0,7%	0,0%	1,4%	0,0%	0,0%	0,0%	0,0%	2,0%	0,0%	0,0%
NUMERO	0,0%	0,0%	0,7%	5,4%	0,0%	0,0%	0,0%	2,7%	0,0%	0,7%
PROBLEMA	0,0%	0,0%	0,0%	0,0%	1,4%	0,7%	0,0%	0,0%	0,0%	0,0%
RESTA	0,0%	0,0%	1,4%	2,0%	0,0%	1,4%	0,0%	0,0%	0,0%	0,0%
SIGNO	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,0%
SUMA	2,0%	0,0%	2,7%	0,0%	0,7%	1,4%	0,0%	1,4%	0,0%	0,0%

Table 3. Percentage of appearance of a word in a sequence of appearance in the survey

	MUNICIPAL	SUBENCIONADO	PARTICULAR
DIVISION	68,0%	56,0%	74,0%
FRACCION	45,0%	58,0%	87,0%
LETRA	56,0%	64,0%	45,0%
MATEMATICA	68,0%	73,0%	67,0%
MULTIPLICACION	89,0%	78,0%	47,0%
NUMERO	67,0%	85,0%	84,0%
PROBLEMA	74,0%	69,0%	34,0%
RESTA	67,0%	56,0%	56,0%
SIGNO	54,0%	67,0%	67,0%
SUMA	64,0%	56,0%	75,0%
Promedio del %	65,2%	66,2%	63,6%

6 Conclusions

This work presented a Bayesian approach to model lexical availability in a certain population regarding a specific center of interest. Our work follows Echeverra's model of lexical availability [3] in which graphs display the relationships among the different words in a certain group's lexicon. The model was tested in a group of Chilean high school students in the field of Algebra, showing a 65% of accuracy in predicting the presence of a certain word.

This proposal has also implications in cognitive semantics, specifically in prototype theory. In particular, this model proves that certain lexical units are more likely to occur in a specific population regarding a center of interest.

Acknowledgements. Research Project - Fondecyt 1120911 Disponibilidad Léxica Matemática en estudiantes de Enseñanza Media y su aplicación en Hipermedios Adaptativos, National Commission for Scientific and Technological Research (CONICYT) of Chile.

References

1. Bashir, A., Khan, L., Awad, M.: Bayesian Networks. In: Wang, J. (ed.) Encyclopedia of Data Warehousing and Mining. Idea Group Publishing, Hershey (2005)
2. Dickson, L., Brown, M., Gibson, O.: El aprendizaje de las matemÃticas. Labor, Madrid (1991)
3. Echeverría, M., Vargas, R., Urzúa, P., Ferreira, R.: Dispografo: una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible. RLA Revista de Lingistica Teórica y Aplicada 46(1), 81–91 (2008)
4. Echeverría, M., Herrera, M., Moreno, P., Pradenas, F.: Disponibilidad léxica en educación media. RLA Revista de Lingistica Teórica y Aplicada 25, 55–115 (1987)
5. Elfman, J.: An alternative view of the mental lexicon. Trends in Cognitive Sciences 8(7), 301–306 (2004)
6. Ferreira, R., Echeverría, M.: Redes semánticas en el léxico dIsponible de inglés L1 e inglés LE. Onomázein 21(1), 133–153 (2010)
7. Giammatteo, M., Albano, H., Basualdo, M.: Competencia léxica y aprendizaje. Interrelación de resultados de investigaciones realizadas en los niveles medio y superior. Filología XXXIII 1-2, 31–57 (2000)
8. Hernández-Muñoz, N., Izura, C., Ellis, A.: Cognitive aspects of lexical availability. European Journal of Cognitive Psychology 18(5), 730–755 (2006)
9. Jurafsky, D.: A probabilistic model of lexical and syntactic access and disambiguation. Cognitive Science 20, 137–194 (1996)
10. McQueen, J., Huettig, F.: Changing only the probability that spoken words will be distorted changes how they are recognized. Journal of the Acoustic Society of America 131(1), 509–517 (2012)
11. Narayanan, S., Jurafsky, D.: Bayesian models of human sentence processing. In: COGSCI-1998, pp. 752–757. Lawrence Erlbaum, Madison (1998)
12. Narayanan, S., Jurafsky, D.: A Bayesian Model Predicts Human Parse Preference and Reading Time in Sentence Processing. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems, vol. 14, pp. 59–65. MIT Press, Cambridge (2002)

13. Orton, A.: DidÃctica de las matemÃticas. MEC y Morata, Madrid (1990)
14. Pearl, J.: Bayesian networks. Department of Statistics, UC Los Angeles (2011), http://escholarship.org/uc/item/5fx179hb (retrieved)
15. Pimm, D.: Speaking Mathematically. Routledge and Kegan Paul, New York (1987)
16. Rotherry, A.: Children reading mathematics. College of Higher Education, Worcester (1980)
17. Salcedo, P., Del Valle, M.: Disponibilidad Léxica Matemática en Estudiantes de Ensenanza Media de Concepción, Chile. In: XIV Evento Internacional Matecompu 2012. La Ensenanza de la Matemática, la Estadística y la Computación, Cuba, Universidad Pedagógica de Cuba, Matanzas (2012)
18. Salcedo, P., Nail, O.: Análisis de relaciones semánticas del léxico disponible en matemáticas en un hipermedio adaptativo. In: TISE 2012, Congreso Internacional de Informática Educativa, Santiago de Chile (Diciembre 2012)
19. Salcedo, P., Ferreira, A., del Valle, M., Nail, O.: Disponibilidad Léxica Matemática en estudiantes de Enseñanza Media y su aplicación en Hipermedios Adaptativos. In: Proyecto Fondecyt 1120911, de la Comisión Nacional de Investigación Científica y Tecnológica de Chile, en ejecución entre los años (2012)
20. Truran, K.: Children's understanding of random generators. Short oral communications. In: Proceeding of the XVIII International Conference for the Psychology of Mathematics Education. University of Lisbon (1994)

Heart Sound Segmentation in Noisy Environments

C. Castro Hoyos, S. Murillo-Rendón, and C.G. Castellanos-Dominguez

Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales
{ccastroh,smurillor,cgcastellanosd}@unal.edu.co

Abstract. This paper presents a methodological framework for the phonocardiographic signal segmentation in telemedicine environments. Framework includes a two-stage noise analysis (high frequency components detection and temporal and spectral periodicity signatures) and heart cycle duration estimation from cyclic frequency spectrograms. Signal boundaries are calculated based on time duration of systolic and diastolic phases. The results are assessed by electrocardiographic based segmentation using the R-peak of heartbeats as signature for delimitation of PCG heart sounds. Approach is tested over a heart sound database with 232 $20-$second recordings. Results show a detection probability of 87% and a false detection probability of 13%.

Keywords: Phonocardiogram, Cycle Frequency, Periodicity.

1 Introduction

Since heart auscultation is a low cost and non-invasive diagnosis method, it becomes a basic diagnostic instrument for cardiac related phenomena. In this regard, heart sounds recorded by auscultation or *phonocardiography* (PCG) are the primary tool for the correct mechanical functional assessment of the heart mechanism. In automatic heart sound analysis, the heart mechanics are assessed by the detection and identification of other sounds within the cardiac cycle that do not correspond to the natural opening and closing of the heart muscle valves. However, this identification requires high medical expertise and its recognition can be affected by the inherent limitations of the human hear for low frequency sounds [1]. Thus, segmenting of the heart sounds is an important task to be carried out. Natural heart sounds are formed by the sounds called S1 and S2, which are originated by the opening and closing valves of the heart muscle, respectively. Between each sound, there is a silence period called systolic and diastolic phase, thereby, systolic interval is conformed by the S1 sound and the systolic phase and diastolic interval by the S2 sound and diastolic phase. Mechanic heart malfunction usually leads to the occurrence of more sounds between S1 and S2, that is, during systolic or diastolic phases. This whole compound of fundamental components forms the heart sound signal over periods called cardiac cycles, and is the basic unit for heart sound analysis.

To the present day, several efforts on developing robust tools have been reported in the segmentation analysis and fundamental sound localization; most of them take advantage of the external measures extracted from cardiac activity (namely, electrocardiogram (EKG) and heart rate variability (HRV)). In [2], for instance, segmentation is carried out using the energy plots and noise detection through a comparative procedure

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 254–263, 2013.

over the fixed length segments with respect to the complete signal energy. However, the timing parameters involved for the heart sound location do not take into consideration HRV and requires the average heart rate previous to segmentation procedure, information that might not be available. In [3], a methodology for detecting noisy PCG signals is proposed that takes into account temporal, spectral and time-frequency periodicity signatures. Yet, the analysis strongly depends on the correct estimation of the heart sound duration estimated from a time sliding window and its singular value decomposition; this methodology can be time-consuming and might not lead to good estimates for short duration signals. To cope with this issue, a dynamic clustering algorithm that does not use EKG information and deals with HRV by means of instantaneous cyclic frequency is reported in [4], however, this procedure requires the PCG signal segment to be decomposed into Gaussian modulated atoms, for which parameters are heuristically fixed and therefore leading to a time-consuming task.

Present study proposes a methodological framework for the heart sound segmentation using high frequency noise detection stages [3] and time-frequency dependent information [5] to assess the quality of the PCG signal. The two-stage noise analysis avoids recordings with strong presence of noise that are further to be segmented, thereby timing location thresholds are used as a basis for heart sound beginning identification. Accurate heart sound boundaries are then set by means to time-dependent heart cycle duration through cycle frequency estimation, as proposed in [4]. The segmentation results are then presented in terms of true and false detection probability, which are assessed employing as ground truth the EKG based segmentation.

2 Materials and Methods

2.1 Estimation of Heart Sound Duration by Removing High Frequency Noise

PCG signals have cyclo-stationarity behavior, that is, the heart sound fundamental components (termed S1-S2) are repeated cyclically as the heart beat does. Therefore, to calculate the time-varying fundamental cycle frequency, the short time cycle frequency spectrum (STCFS) is introduced $\gamma_x(\alpha, t) = R_x(\alpha, t)/R_x(0, t)$, where $R_x(\alpha, t)$ is the cyclic autocorrelation function computed over sliding rectangle windows [5]:

$$R_x(\alpha, t) = \int_{t-\sigma}^{t+\sigma} x(\tau)^2 e^{-j2\pi\alpha\tau} d\tau \ , \tag{1}$$

where α is the cycle frequency and the sliding window expands over $[t - \sigma, t + \sigma]$.

The instantaneous cycle frequency (ICF) is then tracked at the greatest value of $\gamma_x(\alpha, t)$ for a given time instant t in a given frequency range.

$$\delta(t) = \arg\max\left[\gamma_x(\alpha, t)\right] \Big|_{\delta(t^-)-\lambda}^{\delta(t^-)+\lambda} \tag{2}$$

where λ is a threshold parameter to confine the search within a given range and $\delta(t^-)$ is the cycle frequency at earlier time t^-. Thus, the heart sound duration for a given time t_0 is defined as $\tau_d^{(t_0)} = \delta(t_0)^{-1}$.

However, the estimation of cyclic autocorrelation function $R_x(\alpha, t)$ is very sensitive to the noise influence. So, noisy signal removal is to be strongly considered as preprocessing stage. Generally, PCG signals that include spectral components lower that 150 *Hz* can be affected by high frequency external noise and artifacts, which lead to poor analysis and segmentation of heart sounds. Particularly, high frequency external noise shows bursting behavior focusing its influence within very narrow intervals of time, but conveying high energy. Grounded on this fact, the detection of bursting energy segments in signal $\boldsymbol{x}$, being the discrete transformation of PCG signal $x(t)$ of duration T, is suggested in [1] as indicator for removing high frequency noise. Namely, given the energy estimation $e_x[s]$ over signal segment s, it holds that

$$e_x[s] - \tilde{e}_x \geq \eta \tag{3}$$

where η is an a priori fixed threshold parameter, $\tilde{e}_x$ is the average energy, i.e., $\boldsymbol{E}\{e_x[s] : \forall s \in 1, \ldots, N_e\}$, being $N_e = \lfloor N/(lf_s + 1) \rfloor$, with $\lfloor \cdot \rfloor$ the nearest integer approximate of its argument, $N = T/f_s$, and the value l is the considered window analysis of length. Notation $\boldsymbol{E}\{\cdot\}$ stands for the expectation operator.

Although, there are several energy estimators e_x, the Shannon's energy is preferred since it emphasizes the medium intensity sounds and attenuates the effect of low intensity components much more than high intensity components [6]. The Shannon's energy is defined as follows:

$$e_x = \sum_{k=(s-1)l+1}^{sl} -x[k]^2 \log\left(x[k]^2\right) \tag{4}$$

2.2 Segmentation of S1 and S2 Events Based on Periodicity Signatures

For signal boundary delimitation of S1 and S2 events, the wavelet energy envelopes can be employed that clearly produces peaks at the fundamental components of the heart sound signal $S1 - S2$, defined as follows [7,8]:

$$\epsilon(t) = \sum_{\forall a} |C_w(t, a)|^2$$

where $C_w(t, a)$ is the wavelet scalogram of considered signal segment $x(\tau)$, $a \in \mathbb{R}$ is the scale in the continuous wavelet transform. Thus, the prominent peaks of $\epsilon(\tau)$ must be related to the S1 and S2 heart sounds.

For a given set of estimated peaks $\{p_i, \forall i = 1, \ldots, P\}$, where P is the total number of basic components S1-S2 in $x(\tau)$, as well as the phase distances $d_1 = \mathrm{d}(p_1, p_2)$ and $d_2 = \mathrm{d}(p_2, p_3)$, being $\mathrm{d}(\cdot)$ is the Euclidean distance operator, the heart sound segment is said to be synchronic i.e., the first peak does correspond to S1 and the second to S2, if all the next conditions are fulfilled:

- The diastolic phase d_2 must be greater than the systolic d_1.
- $d_1 > 0.2\tau_d$, with τ_d the cardiac cycle duration.
- $d_2 > 0.6\tau_d$.
- $d_2 < 0.8\tau_d$ and $d_1 < 0.8\tau_d$.

Otherwise, the sounds are assumed to be not clearly defined by the energy envelope and additionally phase corrections are to be further provided. Basically, the noise and artifacts cause the sounds to be not properly defined and they can be removed by sequentially using the following two signature periodicity-based algorithms:

Time Periodicity Signature. The similarity among heart sounds can be used for estimating the time periodicity signature. Within a time analysis $2T_\epsilon$, specifically, two neighboring segments, $x_i(t)$ and $x_{i+1}(t)$ (with $t \in [iT_\epsilon, (i+1)T_\epsilon]$) are said to have shape similarity, if an introduced given proximity distance, $\rho_x(x_i, x_{i+1})$, is lower than an a priori given threshold parameter η_ρ, that is:

$$\rho_x(x_i, x_{i+1}) - \eta_\rho \leq 0 \tag{5}$$

As a concrete proximity measure, the radial distance between the prominent peaks (those indicating repetition of a heart sound within the signal segment $2T_\epsilon$,) is suggested in [3], defined as:

$$\rho_x(x_i, x_{i+1}) = \frac{\langle r_x\left(iT_\epsilon, \tau\right), r_x\left((i+1)T_\epsilon, \tau\right)\rangle}{\|r_x\left(iT_\epsilon, \tau\right)\| \, \|r_x\left((i+1)T_\epsilon, \tau\right)\|}, \tag{6}$$

where $i = 1, \ldots, N_h$, being N_h the number of heart sounds occurring within the fixed portions. Notations $\langle\cdot\rangle$ and $|\cdot|$ stand for the inner product and Euclidean norm of their arguments, respectively. Function $r_x\left(\tau\right)$ is the autocorrelation of an energy envelope of $x\left(t\right)$, said $\epsilon_x\left(t\right)$, defined as:

$$r_x\left(T_\epsilon; \tau\right) = \frac{\int_{T_\epsilon} \epsilon_x\left(t\right) h\left(t\right) \epsilon_x\left(t-\tau\right) h\left(t-\tau\right) dt}{\int_{T_\epsilon} h\left(t\right) h\left(t-\tau\right) dt} \tag{7}$$

where $h\left(t\right)$ is a Hamming window. The time analysis T_ϵ is a fixed duration holding at least 3 complete cardiac cycles of the PCG signal, $T_\epsilon \approx 3\tau_d$.

Time-Frequency Periodicity Signature. The heart sound signal concentrates most of energy in the frequency range $0 - 600$ *Hz*, within which the pairwise autocorrelation functions between different bands should exhibit similar behaviour. Moreover, the prominent peaks of the autocorrelated energy occur somehow aligned and the peak width gradually decreases as the bandwidth range increases. Therefore, a given time-frequency periodicity measure, ρ_λ, between two neighboring bands, i and $i+1$, should keep the same (either increasing or decreasing) monotony evolution and any interruption of this monotony serves as and indicator of persistent noise in the signal, i.e.:

$$\operatorname{sgn}(\rho_\lambda(i) - \rho_\lambda(i+1)) = \text{const} \tag{8}$$

The time-frequency PCG representation, $X(\omega, t)$, is divided evenly into 15 frequency bands, and for each one the autocorrelation of the energy envelope (see Eq. (1)) is computed, $r_x^{(k)}(T_\epsilon, \tau)$, with $k = 1, \ldots, 15$. To assess the linear dependency, contiguous ascending bands are grouped together as follows:

$$\boldsymbol{A}^{(i)} = [r_x^{(i)}\left(T_\varepsilon, \tau\right) r_x^{(i+1)}\left(T_\varepsilon, \tau\right) \, \cdots \, r_x^{(i+4)}\left(T_\varepsilon, \tau\right)]^\top$$

from which the singular value set $\{\lambda_j^{(i)}\}$ is computed, where $j-$th indicates the eigenvalue place in descending order and $i-$th is the associated matrix.

So, as time-frequency periodicity measure, the following singular value ratio is calculated:

$$\rho_\lambda(i) = \left(\lambda_2^{(i)}/\lambda_1^{(i)}\right)^2,\ i = 1, 6, 11$$

Lastly, the segment signal is considered as free of noise, when both conditions, given in Eqs (5) and (8), hold. Otherwise, the underlying signal is eliminated.

3 Experimental Setup

3.1 Database and Preprocessing

The used heart sound database for murmur detection belongs to the Signal Processing and Recognition Research Group at the Universidad Nacional de Colombia-Manizales. The database holds 29 PCG recordings for different patients ranging 0 to 20 years old. Each patient contains a $20s$ heart sound signal (with a simultaneous one lead EKG signal) in bell and diaphragm mode for each one of the four traditional auscultation focuses (Aortic, mitral, pulmonary and tricuspid). Signals were recorded at a $44.1kHz$ and 16 bit resolution with an electronic stethoscope and Meditron® software, the acquisition was done in noisy environments as those of health care centers.

PCG signal presents components in a lower frequency range than the original sampling frequency, therefore a $8-$th order Chebychev filter with a downsampling process is implemented with a central cut-off frequency $f_c = 1102.5Hz$, that is a downsampling factor of 20, therefore resulting a final sampling frequency $\hat{f}_s = 2205Hz$. Finally, as recommended in [6], the amplitude of each record was normalized with the absolute maximum of each PCG recording.

3.2 Performance Measure

The simultaneous EKG signal is used to assess the segmentation performance. A traditional EKG segmentation is carried out using the wavelet based QRS detection algorithm given in [9]. Thus, having the proper R-peak locations and provided the correlation between the S1 sound and the QRS complex onset, heart sounds can be picked up and marked. Therefore, the performance evaluation is carried out in terms of true and false detection probability, as discussed in [10]:

$$p_d = \frac{N_D}{N_D + N_M}, \qquad p_f = \frac{N_F}{N_D + N_F}$$

where N_D are the correctly detected points, N_F the number of false points, i.e. a point that is marked but has no correspondence with the true points, and N_M a missed point.

3.3 Experiments

Noise Detection from High Frequency Components. The analysis window for high frequency noise detection is manually set to $l \approx 221$, i.e., the length of the window is approximated to $100ms$. Hence, the number of segments for high frequency noise detection is fixed around 208, η is empirically set to 0.125. As a result, if within 5 continuous segments (0.5 *s* segments of signal) the cumulative energy is higher than η, that signal part is labeled as noisy, and hence the entire signal is also labeled as noisy, otherwise the recording is considered free of high frequency noise.

Heart Sound Duration from Instantaneous Cycle Frequency. The database contains heart rates between 45 and 140 beats per minute corresponding to cycle frequencies from 0.75 to $2.33Hz$. Thus, the maximum cycle frequency is $\alpha_{max} = 2.5Hz$. However, due to digestive and pulmonary noise, artifacts and sounds attenuation, the cyclostationarity of the signals might be affected and the cyclic nature of the signals are reflected on multiples of the fundamental cycle frequency. Thus, this value is set to $\alpha_{max} = 5Hz$ and a search algorithm for the fundamental cycle frequency is required, as detailed in algorithm 1. Then, the sliding window is set in such a way that each signal segment to be correlated (see Eq. (1)) has at least 2 complete cardiac cycles, $\sigma = 2.5\hat{f}_s$, i.e. 2.5 *s* and the search parameter $\lambda = 0.2Hz$, is manually set according to [4].

Noise Detection from Periodicity Signatures. Once $\delta(t)$ is defined, the window length for the temporal periodicity analysis is set to $T_\epsilon = 2s$, then for Eq. (6), $i = 1, \ldots, 5$. Since the pre-detection and heart rate estimation first eliminate highly corrupted signals, the parameter $\eta_\rho = 0.7$ is set higher that the value recommended in [3] to give a more admissible amount of noise. So, the temporal criteria is selected as $\mu_{\rho_x} = \boldsymbol{E}\{\rho_x(x_i, x_{i+1})\}$.

If the time signature criterion holds, the time-frequency signature rule is checked using a spectrogram of each signal portion, computed with a gaussian window. Also, the singular values ratio are computed for all the signal segments. Signals are then labeled as noisy if either $\mu_{\rho_x} < \eta$ or the time-frequency monotony are interrupted.

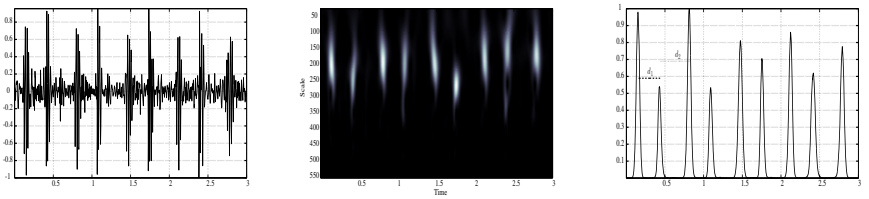
Segments Boundaries and Signal Performance. Finally, the noise-free signal is segmented, taking segments lasting 3 *s* to calculate the C_{Scal} and its respective energy envelope $\epsilon(t)$. The scalogram is computed with a complex Morlet wavelet, and scales a are chosen so the frequency range covers the $25 - 600Hz$ bandwidth. Therefore, none components of the heart sound signal are left out. From Figure 1(a) showing a signal portion and its respective scalogram 1(b), envelogram 1(c), one can see that the durations of d_1 and d_2 clearly defines the systolic and diastolic phase intervals. After obtaining $\epsilon_x(t)$, segmentation is done according to Algorithm 2.

4 Results and Discussion

Figure 2(a) shows an example of high frequency components that are detected within the first noise detection stage (colored as grey areas in the signal). It can be seen how

Algorithm 1. ICF Search Algorithm

Input Signal $x(t), \boldsymbol{\alpha}, \sigma, \lambda$.
Compute $R_x(\alpha, t)$ and $\gamma_x(\alpha, t)$. Get $\delta_0(t)$ as in Eq. (2).
Do $E_{CF} = \sum_{\forall \boldsymbol{t}} R_x^{\alpha}(t)$, the cycle frequency envelope, and calculate its autocorrelation function R_E.
Find the peaks of R_E, P_i. $i = 1, \ldots, n_b$, the number of peaks.
if $n_b = 1$ **then**
  There is only one peak and the ICF is not clearly defined in any frequency band, i.e. noise might be affecting the signal. $h = 0$.
else if $n_b = 2$ **then**
  if $P2 < 0.6$ **then**
    Both peaks are under the minimum for heart cycle frequency.
  else if $P_2 < 2.4$ **then**
    $h = 2$;
  else
    $h = 3$;
  end if
else
  $D(k) = P_{k+1} - P_k \quad \forall k = 2, \ldots, n_b$
  $\mu_{\delta_0} = \boldsymbol{E}\{\delta_0(t)\}$, $\mu_D = \boldsymbol{E}\{D(k)\}$, $h = [\mu_{\delta_0}/\mu_D]_{\text{round off}}$
end if
if $h = 0$ **then**
  The noise is corrupting the ICF, now cycle frequency is achieved within small error.
else
  $\delta(t) = \delta_0(t)/h$.
end if
Return $\delta(t)$.



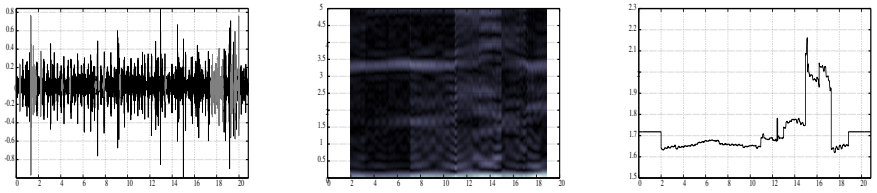
(a) Noise-Free PCG Signal segment (b) Scalogram (c) Energy Envelope

Fig. 1. Scalogram and envelogram of noise-free signal example

Algorithm 2. Segmentation Procedure

Input PCG signal $x(t)$, EKG signal $y(t)$, ICF $\delta(t)$, $t_o = 0, t_f = 3$, set $i = 1$.
Get R-peaks locations of the EKG signal $\boldsymbol{\xi}$.
while $t_f < 20$ **do**
  $x_i = x(\tau), \tau = \{t_o, t_f\}$.
  Get Cycle duration at t_o, $C_i = 1/\delta(t_o)$ and $\epsilon_{x(\tau)}$ and find its prominent peaks.
  Check synchronism conditions as in section 2.2.
  if The conditions hold **then**
    $\kappa_i = p_1 - 0.05\hat{f}_s$
  else
    $\kappa_i = p_2 - 0.075\hat{f}_s$
  end if
  $t_o = \kappa_i + C_i, t_f = t_o + 3$.
end while
Compute the true and false detection probability using $\boldsymbol{\xi}$ and $\boldsymbol{\kappa}$.

the high frequency noise affects computation of the STCFS (Figure 2(b)) as well as the instantaneous cycle frequency (2(c)), exhibiting a high frequency peak of $2.1Hz$ around the 16 s, which is higher that the real average cycle frequency 1.7 measured in the rest of the signal. As a result, from the total number of 232 recordings, 38 are found to have high-frequency noise.



(a) PCG Signal with High Frequency Noise (b) STCFS for a contaminated signal (c) ICF for a contaminated signal

Fig. 2. STCFS and ICF estimation for contaminated signal example

The remaining signals pass through the second noise detection stage, where 65 signals are found to have noise components affecting temporal and time-frequency conditions. As seen from Table 1(b), mitral and pulmonary focuses are the more affected by noise, due to their localization near the lung structure, respiratory noise induces undesired components over the heart sound. The other foci are found to be affected by internal digestive and externals sounds as human voice or crying. The periodicity signature additionally shows that the tricuspid focus has several noisy signals, this may be attributed to the localization distance to the heart valves mechanism; sounds are found to be attenuated in most of these cases, thus causing the time localization of the spectral peaks not matching the expected heart cycle frequency.

Table 1. Number of noisy signals in pre-detection and periodicity signature stages

(a) Pre-detection Noise Summary

	Aortic	Mitral	Pulmonary	Tricuspid
Bell	1	6	7	4
Diaphragm	4	5	7	4

(b) Periodicity Signatures Noise Summary

	Aortic	Mitral	Pulmonary	Tricuspid
Bell	8	6	9	10
Diaphragm	6	10	7	9

Finally, a subset of 129 signals is considered noise-free. The signals are segmented as explained in section 3.3. True and false detection values are calculated for the beginning of the heart sounds and the estimated heart sound durations at those time moments. Tables 2(a) and 2(b) show the average values of p_d and p_f with its respective deviation for the four foci in both diaphragm and bell mode.

Table 2. Segmentation Performance

(a) True Detection Probability

Mode	Aortic	Mitral	Pulmonary	Tricuspid
Bell	0.903 ± 0.091	0.849 ± 0.169	0.878 ± 0.106	0.882 ± 0.079
Diaphragm	0.850 ± 0.126	0.848 ± 0.088	0.855 ± 0.103	0.892 ± 0.095

(b) False Detection Probability

Mode	Aortic	Mitral	Pulmonary	Tricuspid
Bell	0.121 ± 0.085	0.118 ± 0.036	0.107 ± 0.025	0.138 ± 0.073
Diaphragm	0.122 ± 0.03	0.116 ± 0.029	0.185 ± 0.168	0.134 ± 0.069

The cycle duration estimation has an average of 98% accuracy, so the segmentation is only really affected by presence of disruptive peaks in the envelopes due to high amplitude murmurs or some in-band noise that could not be detected. On the other hand, the true detection probability of the heart sounds beginnings has an average value of 87% and an average false detection probability of 13%, showing a robust segmentation performance under considering the poor acquisition conditions.

As seen, aortic and tricuspid signals in Bell and Diaphragm modes have the higher performance, however, aortic focus shows lower deviation due to its far localization from the lung structure. In general, it can be observed the higher deviation and lower performance in the diaphragm auscultation mode, which can be easily explained by the frequency range of operation of the auscultation mode, since diaphragm covers higher frequencies and detects murmurs with higher amplitude that the bell mode, inducing higher false peaks in the energy envelope and therefore errors in the timing conditions.

5 Conclusions and Future Work

A robust heart sound signal segmentation framework is presented as an alternative for noisy environments. Discussed analysis tools, such as high frequency noise detection and periodicity signatures, allow to discriminate signals that are contaminated with the most common noise sources in auscultation. Segmentation shows a steady and constant

performance with a true detection probability of 87% for a total of 129 signals with different heart rates, some including rate variation over the signal.

Noise analysis enables the implementation of low complexity segmentation procedures such as time duration conditions, reducing the computational cost, which is an advantage for telemedicine environments. Additionally, the heart cycle durations estimated from cycle frequency spectrums are more accurate and the introduced band-search procedure allows to supply a more robust analysis against noise.

As future work, it is proposed to develop a more robust and faster segmentation procedure, that does not only use the timing conditions but also spectral and energy content of the heart sounds, so the segmentation uses joint analysis to detect the sound fundamentals (S1-S2) and other components related to murmurs.

Acknowledgments. This work is supported by the "*Aprendizaje de Máquina a partir de mútiples expertos en clasificación multiclase de señales de voz*" and "*Discrimination of ECG Signals via Time-Frequency Representations*" projects associated with the "Jóvenes Investigadores e Innovadores 2011" program funded by COLCIENCIAS and the project "*Servicio de Monitoreo Remoto de Actividad Cardíaca para el Tamizaje Clínico en la Red de Telemedicina del Departamento de Caldas*" financed by "convocatoria conjunta entre Universidad Nacional-Manizales Universidad de Caldas".

References

1. Ari, S., Kumar, P., Saha, G.: A robust heart sound segmentation algorithm for commonly occurring heart valve diseases. Journal of Medical Engineering and Technology 32(6), 456–465 (2008)
2. Ari, S., Saha, G.: On a robust algorithm for heart sound segmentation. Journal of Mechanics in Medicine and Biology 7(2), 129–150 (2007)
3. Kumar, D., Carvalho, P., Antunes, M., Paiva, R.P., Henriques, J.: Noise detection during heart sound recording using periodicity signatures
4. Tang, H., Li, T., Qiu, T., Park, Y.: Segmentation of heart sounds based on dynamic clustering. Biomedical Signal Processing and Control 7(5), 509–516 (2012)
5. Tang, H., Li, T., Qiu, T.: Cardiac cycle detection for heart sound signal based on instantaneous cycle frequency. In: 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), vol. 2, p. 676 (October 2011)
6. Delgado-Trejos, E., Quiceno-Manrique, A.F., Godino-Llorente, J.I., Blanco-Velasco, M., Castellanos-Dominguez, G.: Digital Auscultation Analysis for Heart Murmur Detection. Annals of Biomedical Engineering
7. Ergen, B., Tatar, Y., Gulcur, H.O.: Time-frequency analysis of phonocardiogram signals using wavelet transform: a comparative study. Comput. Methods Biomech. Biomed. Engin. 15(4), 371–381 (2012)
8. Choi, S., Jiang, Z.: Comparison of envelope extraction algorithms for cardiac sound signal segmentation. Expert Syst. Appl. 34(2), 1056–1069 (2008)
9. Kohler, B.-U., Hennig, C., Orglmeister, R.: The principles of software qrs detection. IEEE Engineering in Medicine and Biology Magazine 21(1), 42–57 (2002)
10. Sörnmo, L., Laguna, P.: Bioelectrical Signal Processing in Cardiac and Neurological Applications. Academic Press Series in Biomedical Engineering. Academic Press (2005)

Kernel Spectral Clustering for Motion Tracking: A First Approach

D. Peluffo-Ordóñez*, S. García-Vega, and C.G. Castellanos-Domínguez

Universidad Nacional de Colombia – Manizales,
Km 7 vía al Magdalena
{dhpeluffoo,segarciave,cgcastellanosd}@unal.edu.com
http://www.fia.unal.edu.co/gta/signal/index.php

Abstract. This work introduces a first approach to track moving-samples or frames matching each sample to a single meaningful value. This is done by combining the kernel spectral clustering with a feature relevance procedure that is extended to rank the frames in order to track the dynamic behavior along a frame sequence. We pose an optimization problem to determine the tracking vector, which is solved by the eigenvectors given by the clustering method. Unsupervised approaches are preferred since, for motion tracking applications, labeling is unavailable in practice. For experiments, two databases are considered: Motion Caption and an artificial three-moving Gaussian in which the mean changes per frame. Proposed clustering is compared with kernel K-means and Min-Cuts by using normalized mutual information and adjusted random index as metrics. Results are promising showing clearly that there exists a direct relationship between the proposed tracking vector and the dynamic behavior.

Keywords: Kernels, motion tracking, spectral clustering.

1 Introduction

Nowadays, dynamic or time-varying data analysis is of great interest for the scientific community, specially in automation and pattern recognition. Video analysis [1] and motion identification [2] are some of the applications. Spectral matrix analysis is one of the approaches to address this issue. Spectral techniques, particularly within the graph-cut framework, has been a suitable tool in several aspects of interest in pattern recognition and machine learning even when data are time-varying, such as the estimation of the number of clusters [3], feature extraction and selection [4] as well as clustering [5] and classification [2]. Regarding spectral clustering, in [6] it is introduced the so-called kernel spectral clustering (KSC), which is based on least-square support vector machines and has shown to be a powerful tool for clustering hardly separable data allowing

* This work is supported by *"Jóvenes Investigadores"* COLCIENCIAS program with the project entitled *"Comparativo de métodos kernel para agrupamiento espectral de datos desde un enfoque primal-dual"*.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 264–273, 2013.

also extensions to out-of-samples [7]. In another work [4], the authors present a definition of feature relevance aimed to selecting a subset of features founded on spectral properties of the Laplacian of data matrix. This approach is based on a continuous ranking of the features by means of a least-squares maximization problem. Such maximization is done over a functional such that sparse solutions for the ranking values are obtained by a spectral decomposition of non-negative matrix [4]. What makes this approach interesting within the context that we are concerned in this work is the possibility to get a ranking value for each frame in the process of time-varying data analysis. In addition, the feature relevance approach measures the relevance of a subset of features against its influence on the cluster assignment.

In this work, from the KSC combined with the relevance ranking proposed in [4], we introduce a novel approach to track moving samples or frames matching each sample to an unique meaningful value. Notice that we prefer the unsupervised approaches since, in practice, labeling is not available for motion tracking applications. Our approach works as follows: it starts by clustering the input data by means of KSC with a manually established of groups and a determined kernel function. Then, we take the kernel matrix as another data representation that can be linearly projected in order to apply the relevance ranking over the samples instead of the features as proposed in [4]. The projection matrix is obtained as a sparse solution of a quadratic optimization problem, where a projection energy (also called variance) term is maximized. It is proved that projection matrix is the same as the eigenvector matrix given by KSC. Finally, a tracking vector is obtained by a linear combination of such eigenvectors similarly as described in [8]. Experiments are carried out using two databases: Motion Caption and an artificial three-moving Gaussian in which the mean is changed throughout the frames. Proposed KSC is compared with kernel K-means and Min-Cuts [9]. To quantify the clustering performance, normalized mutual information [10] and adjusted random index [11] metrics are used. Obtained results are promising showing clearly that there exists a direct relationship between the proposed tracking vector and the analyzed dynamic data.

2 Theoretical Background

2.1 Kernel Spectral Clustering

The aim of kernel spectral clustering (KSC), introduced in [6], is to divide a data set of N samples into K homogeneous and disjoint subsets. Consider that the data are arranged in the data matrix $\boldsymbol{X} \in \mathbb{R}^{N \times d}$ assuming that each sample vector $\boldsymbol{x}_i$ is d-dimensional, then $\boldsymbol{X} = [\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_N^\top]^\top$. Assume a latent variable model for the projections $\boldsymbol{E} \in \mathbb{R}^{N \times n_e}$ as $\boldsymbol{E} = \boldsymbol{\Phi W} + \mathbf{1}_N \otimes \boldsymbol{b}^\top$, being matrix $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}(\boldsymbol{x}_1)^\top, \ldots, \boldsymbol{\phi}(\boldsymbol{x}_N)^\top\right]^\top$, $\boldsymbol{\Phi} \in \mathbb{R}^{N \times d_h}$, is a high dimensional representation of data such that $\phi(\cdot)$ maps data from the original dimension to a higher one d_h, i.e., $\boldsymbol{\phi}(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d_h}$, $\boldsymbol{W} = [\boldsymbol{w}^{(1)}, \cdots, \boldsymbol{w}^{(n_e)}]$ is the weighting factor matrix, $\boldsymbol{W} \in \mathbb{R}^{d_h \times n_e}$, and $\boldsymbol{b} = [b_1, \ldots, b_{n_e}]$ is a vector containing bias terms, $\boldsymbol{b} \in \mathbb{R}^{n_e}$. Notation

$\otimes$ the Kronecker product and term n_e denotes the number of considered support vectors. Then, according to [6] within a least-square-support vector machine framework, a matrix primal formulation of KSC can be stated as:

$$\min_{\boldsymbol{E},\boldsymbol{W},\boldsymbol{b}} \quad \frac{1}{2N}\,\mathrm{tr}(\boldsymbol{E}^\top \boldsymbol{V}\boldsymbol{E}\boldsymbol{\Gamma}) - \frac{1}{2}\,\mathrm{tr}(\boldsymbol{W}^\top\boldsymbol{W}); \quad \text{s.t.} \quad \boldsymbol{E} = \boldsymbol{\Phi}\boldsymbol{W} + \mathbf{1}_N \otimes \boldsymbol{b}^\top \tag{1}$$

where $\boldsymbol{\Gamma} = \mathrm{Diag}([\gamma_1, \ldots, \gamma_{n_e}])$ is a diagonal matrix formed by the regularization terms. Notations $\mathrm{tr}(\cdot)$ and $\otimes$ denote the trace and the Kronecker product, respectively.

To solve the KSC problem, we form the corresponding Lagrangian of previous problem, as follows:

$$\mathcal{L}(\boldsymbol{E},\boldsymbol{W},\boldsymbol{\Gamma},\boldsymbol{A}) = \frac{1}{2N}\,\mathrm{tr}(\boldsymbol{E}^\top\boldsymbol{V}\boldsymbol{E}) - \frac{1}{2}\,\mathrm{tr}(\boldsymbol{W}^\top\boldsymbol{W}) - \mathrm{tr}(\boldsymbol{A}^\top(\boldsymbol{E} - \boldsymbol{\Phi}\boldsymbol{W} - \mathbf{1}_N \otimes \boldsymbol{b}^\top))$$

where matrix $\boldsymbol{A} \in \mathbb{R}^{N\times n_e}$ contains the Lagrange multiplier vectors such that $\boldsymbol{A} = [\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(n_e)}]$, and $\boldsymbol{\alpha}^{(l)} \in \mathbb{R}^N$ is the l-th vector of Lagrange multipliers.

Then, we determine the Karush-Kuhn-Tucker conditions by solving the partial derivatives on $\mathcal{L}(\boldsymbol{E},\boldsymbol{W},\boldsymbol{\Gamma},\boldsymbol{A})$. Afterwards, by eliminating the primal variables, the optimization problem posed in equation (1) is reduced to the following dual problem: $\boldsymbol{A}\boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{\Phi}^\top\boldsymbol{A}$, where $\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_{n_e})$ is a diagonal matrix formed by the eigenvalues $\lambda_l = N/\gamma_l$, matrix $\boldsymbol{H} \in \mathbb{R}^{N\times N}$ is the centering matrix that is defined as $\boldsymbol{H} = \boldsymbol{I}_N - 1/(\mathbf{1}_N^\top\boldsymbol{V}\mathbf{1}_N)\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{V}$, being $\boldsymbol{I}_N$ a N-dimensional identity matrix. In addition, by applying the kernel trick in such a way that $\boldsymbol{\Omega} \in \mathbb{R}^{N\times N}$ be the kernel matrix $\boldsymbol{\Omega} = [\Omega_{ij}] = \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j), i, j \in [N]$, we have that $\boldsymbol{\Omega} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\top$. Notation $\mathcal{K}(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ stands for the kernel function. Notice that matrix $\boldsymbol{A}$ becomes the eigenvectors. As a result, the set of projections can be calculated as follows:

$$\boldsymbol{E} = \boldsymbol{\Omega}\boldsymbol{A} + \mathbf{1}_N \otimes \boldsymbol{b}^\top \tag{2}$$

Taking into account that the kernel matrix represents the similarity matrix of a graph with K connected subgraphs and assuming $\boldsymbol{V} = \boldsymbol{D}^{-1}$ being $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ the degree matrix defined as $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{\Omega}\mathbf{1}_N)$; we can infer that the $K-1$ eigenvectors associated to the largest eigenvalues are cluster indicators [7]. Therefore, value n_e is fixed to be $K-1$. Afterwards, since each cluster is represented by a single coordinate in the $K-1$-dimensional eigenspace, we can encode the eigenvectors considering that two points are in the same cluster if they are in the same orthant in the corresponding eigenspace [7]. Therefore, by binaryzing the rows of the projection matrix $\boldsymbol{E}$, we obtain the code book as $\widetilde{\boldsymbol{E}} = \mathrm{sgn}(\boldsymbol{E})$, where $\mathrm{sgn}(\cdot)$ is the sign function. Thus, its corresponding rows are codewords, which allow to form the clusters according to the minimal Hamming distance.

2.2 Tracking by KSC

Similarly as the relevance analysis explained in [4] in which a functional regarding a non-negative matrix is introduced, we pose an optimization problem with the

difference that we are concerned of obtaining the ranking values for samples instead of features, as follows: Focusing on the task of interest, consider that the non-negative matrix is chosen as $\boldsymbol{\Omega}$ and the data matrix $\boldsymbol{X}$ is formed in such a way that each row is a frame, i.e., $\boldsymbol{x}_i$ is the vectorization of coordinates representing the i-th frame. By recalling equation (2), an energy maximization problem can be written as:

$$\max_{\boldsymbol{U}} \ \mathrm{tr}(\boldsymbol{U}^\top \boldsymbol{\Omega}^\top \boldsymbol{\Omega} \boldsymbol{U}); \quad \text{s.t.} \quad \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_{n_e} \tag{3}$$

The orthonormal rotation matrix is $\boldsymbol{U}$ is in size $N \times n_e$, such that the linear transformation $\boldsymbol{Z} \in \mathbb{R}^{N \times n_e}$ of kernel matrix is in the form $\boldsymbol{Z} = \boldsymbol{\Omega U}$. Considering the procedure described in section 2.1, we can notice that $\mathrm{tr}(\boldsymbol{U}^\top \boldsymbol{\Omega}^\top \boldsymbol{\Omega U}) = \mathrm{tr}(\boldsymbol{\Lambda}^2)$ and therefore a feasible solution of the problem is $\boldsymbol{U} = \boldsymbol{A}$. Now, as explained in [8], a ranking vector $\boldsymbol{\eta} \in \mathbb{R}^N$ can be expressed as a linear combination of vectors $\boldsymbol{\alpha}^{(l)}$ so:

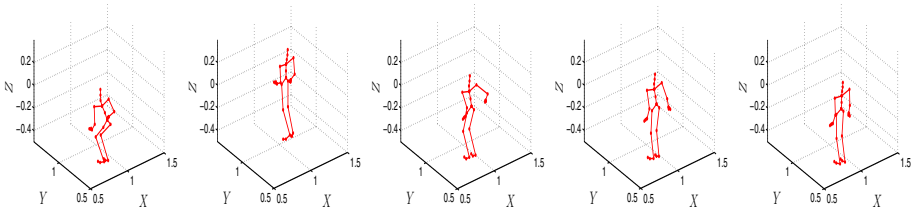
$$\boldsymbol{\eta} = \sum_{l=1}^{n_e} \lambda_l |\boldsymbol{\alpha}^{(l)}| \tag{4}$$

where $|\cdot|$ denotes absolute value. Accordingly, the ranking factor η_i is a single value representing an unique frame in a sequence. In this sense, $\boldsymbol{\eta}$ can be called tracking vector.

3 Experimental Set-Up

3.1 Databases

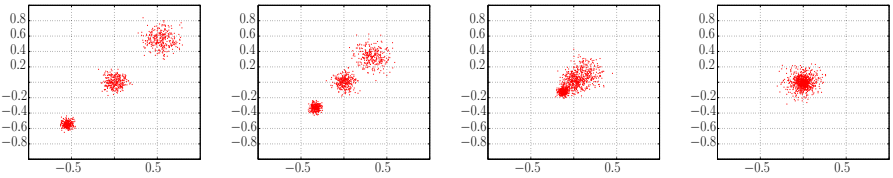
Motion Caption. *The data used in this work was obtained from mocap.cs.cmu.edu. The database was created with funding from NSF EIA-0196217.* In this work, we use the trial number 4 (02_04), particularly, the *subject #2* (jump, balance). Each frame is in size 38×3 representing the coordinates X, Y and Z of the subject's body points, therefore each $\boldsymbol{x}_i$ is 114-dimensional. Since 484 frames are considered, data matrix is in size 484×114.



(a) Frame 95 (b) Frame 140 (c) Frame 182 (d) Frame 280 (e) Frame 404

Fig. 1. Motion Caption Database

Three-Moving Gaussian. An artificial three dimensional Gaussian data sequence is also considered, which consist of Gaussian data with 3 clusters in such a way the deviation standard is static for all the frames and means are decreasing to move per frame each cluster towards each other. Namely, for a total of N_{fm} frames the mean and standard deviation vectors at r-th frame are respectively in the form $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3] = [-20(1 - r/N_{fm}), 0, -20(1 - r/N_{fm})]$ and $\boldsymbol{s} = [s_1, s_2, s_3] = [1, 2, 3]$, being μ_j and s_j the mean and standard deviation corresponding to the j-th cluster, respectively; as well as $j \in \{1, 2, 3\}$ and $r \in \{1, \ldots, N_{fm}\}$. The number of data samples per cluster is 300 and the considered total of frames is 100. Thus, each frame is in size 900×3 which means that $\boldsymbol{x}_i$ is of length 2700 as well as data matrix is $\boldsymbol{X} \in \mathbb{R}^{100 \times 2700}$. In Fig. 2, some frames of moving Gaussian are depicted.



(a) `Frame 19` (b) `Frame 51` (c) `Frame 82` (d) `Frame 100`

Fig. 2. Three Gaussian Database

Data matrices from the both above databases are z-score normalized regarding their columns before starting the clustering process.

3.2 Clustering and Kernel Parameters

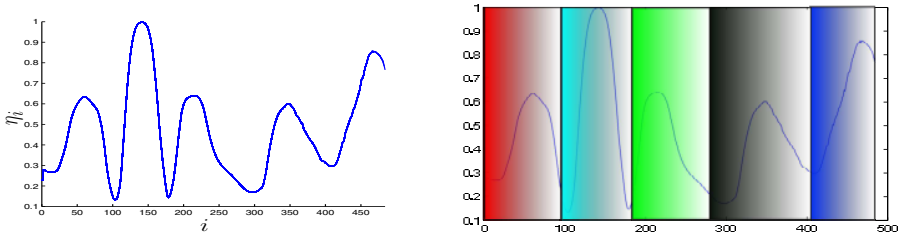
All the experiments are performed under specific initial parameters, namely, the number of clusters K and kernel function. For Motion Caption database (*subject #2*), parameter K is set to be 5. In case of moving Gaussian, we empirically determine that $K = 4$. The considered kernel for both cases is the local-scaled Gaussian kernel [12] defined as: $\Omega_{ij} = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2/(\sigma_i \sigma_j))$, where $\|\cdot\|$ denotes the Euclidean norma and the scale parameter σ_i is chosen as the Euclidean distance between the sample $\boldsymbol{x}_i$ and its corresponding m-th nearest neighbor. Free parameter m is empirically set by varying it within an interval and then it is chosen as that one achieving greatest Fisher's criterion value. In the case of Motion Caption, we obtain $m = 35$; while $m = 10$ for the moving Gaussian.

For comparison purposes, kernel K-means (KKM) and Min-Cuts are also considered [9]. The clustering performance is quantified by two metrics: normalized mutual information (NMI) [10] and adjusted random index (ARI) [11]. Both metrics return values ranged into the interval $[0, 1]$, being closer than 1 when better is the clustering performance.

4 Results

4.1 Results for Motion Caption Database

After applying KSC over matrix $\boldsymbol{X}$, vector $\boldsymbol{\eta}$ is calculated using (4). Fig. 3 is the dynamic analysis achieved by KSC. The tracking vector plotting is shown in Fig. 3(a). As it can be appreciated, $\boldsymbol{\eta}$ has a multi-modal shape. Since the eigenvectors $\boldsymbol{\alpha}^{(l)}$ point out the direction where samples have the most variability measured in term of a generalized inner product ($\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi}$), we can infer that each mode might represent a different cluster. In Fig. 3(b), the reference labeling vector is shown which is obtained by detecting the local minima through a simple search. Such vector is henceforth considered as the reference labels to quantify the clustering performance by NMI and ARI.

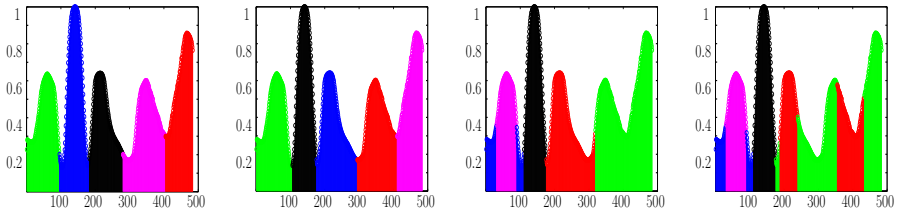


(a) Tracking vector $\boldsymbol{\eta}$ (b) Reference labels

Fig. 3. Dynamic analysis of *Subject # 2* by KSC

In Fig. 4, we can observe the dynamic behavior tracking done for all considered methods. We match the tracking vector with cluster indicators yielded by each method. It is easy to appreciate that KSC performs a better clustering in comparison to the other methods. This fact can be attributed to the coherence and evident linkage between the tracking vector and KSC. Also, note that the samples are organized in sequence, then we expect resultant clusters contain no isolated or disconnected (out of sequence) samples. In other words, clustering indicators reach low values when samples or frames are intermittently clustered along time. Then, KKM and Min-Cuts generate partially wrong grouping. Nevertheless, it is worth mentioning that this experimental framework was done with a specific type of kernel as well as a determined parameter tuning, i.e., maximal Fisher's criterion in this case. Thus, it is possible that such a tuning process may not be appropriate for those methods. Albeit, it must be taken into account that Fisher's criterion is applied regardless the method, that is to say that tuning process is done considering no the clustering method but the nature of data by means of the within- and between-variance ratio (Fisher's criterion). The trial 4 (02_04) is a vertical jump as shown. Then, by comparing Fig. 1 with Fig. 3, we can appreciate that the identified local minima determine five clusters representing five dynamic instances along time, associated to subject #2.

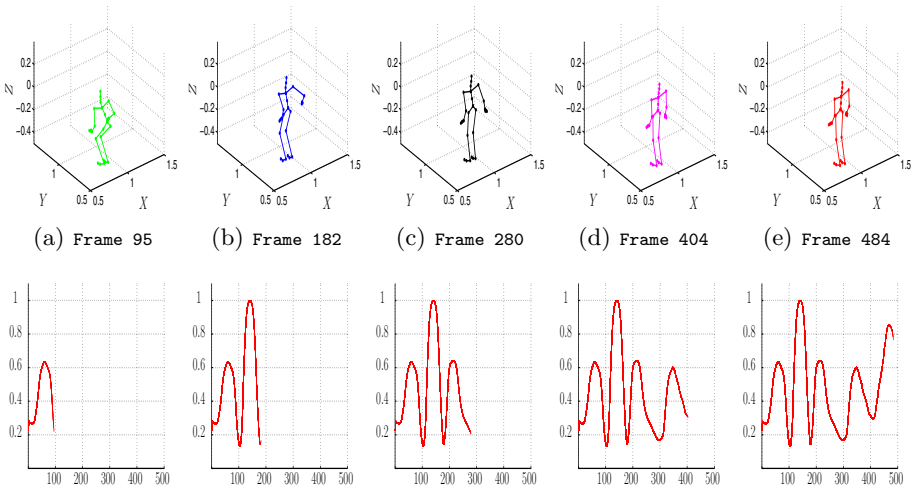
We name those clusters as follows: *starting jump* (between frames 1 and 95), *jumping on the air* (between frames 96 and 182), *arrival to the ground* (between frames 183 and 280), *back step* (between 281 and 404) and *quiet – Standing* (between frames 405 and 484).



(a) `Reference Labels` (b) KSC (c) KKM (d) Min-Cuts

Fig. 4. Clustering results for *Subject # 2*

The direct relationship between tracking vector $\boldsymbol{\eta}$ and the partition of natural movements from *Subject #2* can be plainly appreciated in Fig. 5, where the top row shows one representative frame per cluster (in different color) while bottom row depicts the $\boldsymbol{\eta}$ curve until the last frame of the corresponding cluster.



(a) `Frame 95` (b) `Frame 182` (c) `Frame 280` (d) `Frame 404` (e) `Frame 484`

(f) *starting jump* (g) *jumping - on the air* (h) *arrival to the ground* (i) *back step* (j) *quiet – standing*

Fig. 5. *Subject #2* tracking

As expected, each mode – frames between inflections forming a concave curve – from $\boldsymbol{\eta}$ plotting corresponds to a natural cluster, which may even be determined by simple inspection. The tracking effect can also be verified in Table 1. In terms of the considered measures, we can observe that KSC outperforms both KKM and Min-Cuts, which means that the clustering method might be coherent with the tracking procedure. In this connection, KSC is the most suitable since $\boldsymbol{\eta}$ comes from the eigenvectors.

Table 1. Clustering performance for *Subject # 2* in terms of considered metrics

Measure	Clustering Method		
	KSC	KKM	Min Cuts
NMI	0.8410	0.6982	0.6974
ARI	0.8163	0.5752	0.5400

4.2 Results for Three-Moving Gaussian Database

Similarly to the previous database, now we explore the dynamic behavior when Gaussian data are moving towards each other. Fig. 6 depicts the tracking vectors when varying the number of clusters from 2 to 6 on top, and the corresponding reference labeling vectors at bottom.

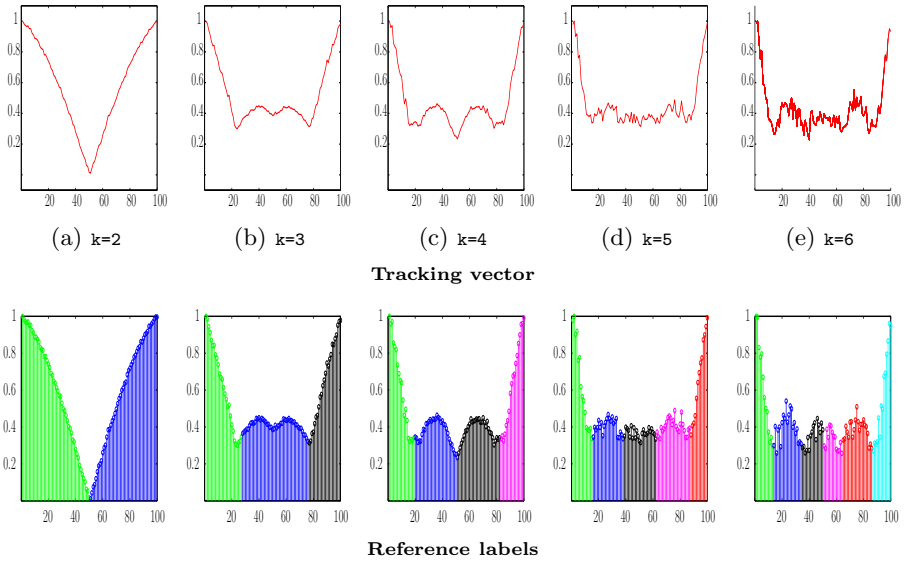
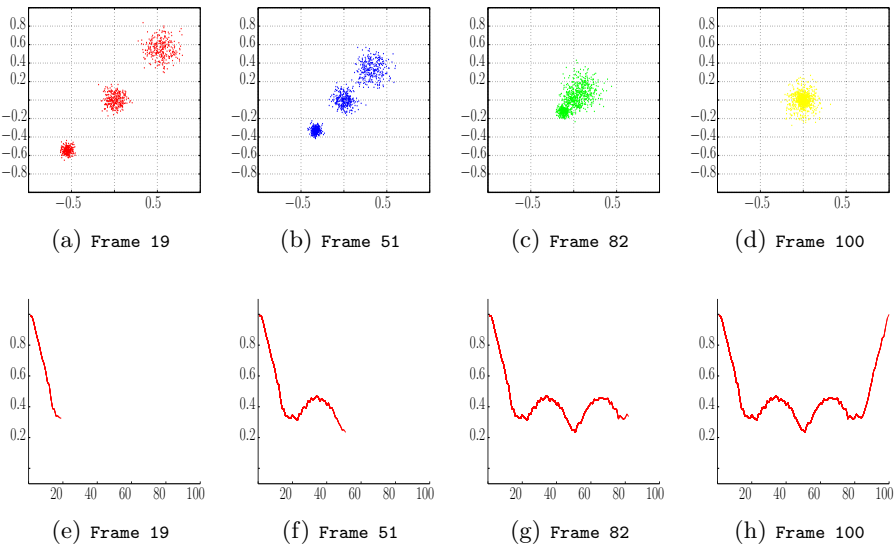


Fig. 6. Tracking Vectors for three-Gaussian moving

Table 2. Effect of number of groups over the *Subject #2*

Measure	K	Clustering Method		
		KSC	**KKM**	**Min-Cuts**
NMI	2	1	0.9291	1
	3	0.5817	0.8068	0.7257
	4	0.8839	0.8306	0.6408
	5	0.7426	0.9470	0.5966
	6	0.6998	0.7803	0.5900
ARI	2	1	0.9600	1
	3	0.4017	0.8135	0.7046
	4	0.8924	0.8227	0.5037
	5	0.5840	0.9416	0.4148
	6	0.5281	0.6761	0.3393



(a) `Frame 19` (b) `Frame 51` (c) `Frame 82` (d) `Frame 100`

(e) `Frame 19` (f) `Frame 51` (g) `Frame 82` (h) `Frame 100`

Fig. 7. Three-moving Gaussian regarding tracking-generated reference labels

As it can be noted, either $K = 2$ or $K = 4$ seems to be the proper number of clusters because in those modes are well formed and then easily identifiable. Then, our approach can also be used to determine the number of groups in a sequence of frames. Indeed, in Table 2 the best performance is reached when clustering with those values of K.

For instance in Fig. 7, the tracking effect with $K = 4$ is shown. Again, each mode represents a single cluster.

5 Conclusions

This work presents a novel approach for motion tracking from a sequence of moving samples or frames, which takes advantage of spectral properties from a kernel-based clustering as well as a relevance ranking procedure. Tracking is done by finding an unique value representing adequately each single frame. Our approach determines a tracking vector that has a direct relationship with the underlying dynamic behavior of analyzed sequence, allowing even to estimate the number of groups as well as the ground truth. As a future work, with the aim of designing an accurate tracking method able to be extended to prediction problems, new spectral properties and techniques are to be explored.

References

1. Shirazi, S., Harandi, M.T., Sanderson, C., Alavi, A., Lovell, B.C.: Clustering on grassmann manifolds via kernel embedding with application to action analysis. In: Proc. IEEE International Conference on Image Processing (2012)
2. Sudha, L., Bhavani, R.: Performance comparison of svm and knn in automatic classification of human gait patterns. Int. J. Comput 6(1), 19–28 (2012)
3. Lee, S., Hayes, M.: Properties of the singular value decomposition for efficient data clustering. IEEE, Signal Processing Letters 11(11), 862–866 (2004)
4. Wolf, L., Shashua, A.: Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach. J. Mach. Learn. Res. 6, 1855–1887 (2005), http://portal.acm.org/citation.cfm?id=1046920.1194906
5. Alzate, C., Suykens, J.: Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA. IEEE Transactions on Pattern Analysis and Machine Intelligence, 335–347 (2008)
6. Alzate, C., Suykens, J.: A weighted kernel PCA formulation with out-of-sample extensions for spectral clustering methods. In: International Joint Conference on Neural Networks, IJCNN 2006, pp. 138–144. IEEE (2006)
7. Alzate, C., Suykens, J.A.K.: Multiway spectral clustering with out-of-sample extensions through weighted kernel pca. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(2), 335–347 (2010)
8. Molina-Giraldo, S., Álvarez-Meza, A.M., Peluffo-Ordoñez, D.H., Castellanos-Domínguez, G.: Image segmentation based on multi-kernel learning and feature relevance analysis. In: Pavón, J., Duque-Méndez, N.D., Fuentes-Fernández, R. (eds.) IBERAMIA 2012. LNCS, vol. 7637, pp. 501–510. Springer, Heidelberg (2012)
9. Guo, C., Zheng, S., Xie, Y., Hao, W.: A survey on spectral clustering. In: World Automation Congress (WAC), pp. 53–56. IEEE (2012)
10. Strehl, A., Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research 3, 583–617 (2002)
11. Hubert, L., Arabie, P.: Comparing partitions. Journal of Classification 1(2), 193–218 (1985)
12. Zelnik-manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems, vol. 17, pp. 1601–1608. MIT Press (2004)

Multi-labeler Analysis for Bi-class Problems Based on Soft-Margin Support Vector Machines

S. Murillo-Rendón[1], D. Peluffo-Ordóñez[1], J.D. Arias-Londoño[2], and C.G. Castellanos-Domínguez[1]

[1] Universidad Nacional de Colombia – Manizales, Caldas, Colombia
[2] Universidad de Antioquia, Medellín, Colombia

Abstract. This work presents an approach to quantify the quality of panelist's labeling by means of a soft-margin support vector machine formulation for a bi-class classifier, which is extended to multi-labeler analysis. This approach starts with a formulation of an objective function to determine a suitable hyperplane of decision for classification tasks. Then, this formulation is expressed in a soft-margin form by introducing some slack variables. Finally, we determine penalty factors for each panelist. To this end, a panelist's effect term is incorporated in the primal soft-margin problem. Such problem is solved by deriving a dual formulation as a quadratic programming problem. For experiments, the well-known Iris database is employed by simulating multiple artificial labels. The obtained penalty factors are compared with standard supervised measures calculated from confusion matrix. The results show that penalty factors are related to the nature of data, allowing to properly quantify the concordance among panelists.

Keywords: Bi-class classifier, multi-labeler analysis, quadratic programming, support vector machines.

1 Introduction

In several supervised pattern recognition problems, a ground truth is beforehand known to carry out a training process. Nonetheless, there are cases where such ground truth is not unique. For instance, in medical environments, the diagnostic judgment given by only one doctor (labeler, panelist, teacher, evaluator) might not be enough since the labeling is greatly related to the panelist's subjectivity and criterion [1]. Another example is the labeling carried out through internet web servers, where a set of labeler are asked to qualify data from different sources aiming to classify them in a determined group. However, because the evaluation is not restricted, that is to say, some non-experts labelers can also label the analyzed data, the additional problem of presence of noisy labels must be considered [2]. Some works have been concerned about this issue. In [3], authors consider a set of experts to determine the crater distribution in Venus surface, by comparing human and algorithmic performance as opposed to simply comparing humans to each other. Moreover, the multi-labeler approach is only the labels average. Other studies, are focused on building proper decision boundaries from multiple-experts labels, but requiring some prior information, [4]. Finally, in [5], the multi-experts task is addressed by a support vector machine (SVM) scheme yielding a suitable approach to

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 274–282, 2013.

penalize panelist mistakes, this methodology gives a good classifier from a set of teachers, compensating the possible labeling errors by evaluating the amount and disposition of support vectors provided for each teacher to the decision boundary.

This work proposes a methodology to quantify the panelist's labeling from a soft-margin support vector machine approach (SMSVM), as a variation to that proposed in [5], the main difference in our optimization problem formulation is done within a quadratic programming framework for estimating penalty factors for labelers instead of a decision function for classification. Moreover, the proposed approach also allows to get a suitable classifier with multi-labeler training, which is further used to determine an estimation of ground truth. Also, it allows for obtaining penalization factors, one for each panelist. Such factors keep the relation between labelers and the structure of data, making evident the labeling vector quality for each labeler.

The outline of this paper is as follows: In section 2, we briefly describe our method to analyze the reliability of panelist labeling. Section 4 shows and discuss the obtained results. Finally, in section 5, some final remarks and conclusions are presented.

2 Methods

The present work proposes a variation of a SM-SVM traditional formulation is introduced, which consists of adding a penalty term and a quadratic constrain in the functional of primal formulation. Such penalty term is aimed to penalize the supposed wrong-labeled data, by means of a linear combination of some penalty factors quantifying the labeler's performance. With this variation, our approach also generates a suitable decision function allowing to obtain an estimation of ground truth, even when the labelers are wrong. In summary, we formulate an optimization problem by generalizing the classifier taking into account different labeling vectors and adding penalty factors.

2.1 Soft Margin Binary SVM Formulation

Let us define the ordered pair $\{\boldsymbol{x}_i, y_i\}$ as the i-th sample where $\boldsymbol{x}_i \in \mathbb{R}^d$ is a d-dimensional feature vector and y_i is its binary class indicator. In this case, $y_i \in \{1, -1\}$. In matrix terms, $\boldsymbol{X} \in \mathbb{R}^{m \times d}$ and $\boldsymbol{y} \in \mathbb{R}^m$, are respectively the data matrix and labeling vector, being d the number of considered features and m the number of samples. We assume an hyperplane model in the form: $\boldsymbol{w} \cdot \boldsymbol{x} + b = \boldsymbol{w}^\top \boldsymbol{x} + b = 0$, where $\boldsymbol{w}$ is an orthogonal vector to the hyperplane, b is a bias term.

Intuitively, for a two-class problem we can establish $f(\boldsymbol{x}) = \operatorname{sign}(\boldsymbol{w}^\top \boldsymbol{x} + b)$ as the decision function. In order to avoid that data point lie in a region where there exists ambiguity to take the decision, we assure that the distance between the hyperplane and any data point must be equal or more than a priori fixed value (i.e,. one value is chosen) to satisfy the condition: $y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b)) \geq 1 \quad \forall i$. Then, the distance between any data point x_i and the hyperplane $(\boldsymbol{w}, b)$ can be calculated as: $\mathrm{d}\left((\boldsymbol{w}, b), \boldsymbol{x}_i\right) = y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b)/||\boldsymbol{w}||_1^2 \geq 1/||\boldsymbol{w}||_2^2$, where $||\cdot||_2$ stands for Euclidean norm. Therefore, we expect that $y_i \simeq \boldsymbol{w}^\top \boldsymbol{x}_i + b$, since upper boundary is $1/||\boldsymbol{w}||^2$. Then, the classifier objective function to be maximized can be written as:

$$\min_{\boldsymbol{w}} \frac{1}{2}||\boldsymbol{w}||^2, \quad \text{s.t.} \quad y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) = 1; \ \ \forall i \tag{1}$$

By relaxing (1) using a quadratic constrain, we can write the following SVM-based formulation:

$$\min_{\boldsymbol{w}} f(\boldsymbol{w}|\lambda, b) = \min_{\boldsymbol{w}} \frac{\lambda}{2}||\boldsymbol{w}||^2 + \frac{1}{m}\sum_{i=1}^{m}(1 - \boldsymbol{y}_i(\boldsymbol{w}^\top \boldsymbol{x} + b))^2 \tag{2}$$

where λ is a regularization parameter. Previous formulation is a hard margin approach, i.e., data points are not expected to lie on the decision function boundary. Recalling (2), we can extend the functional to a soft margin formulation incorporating a slack variable ξ_i, such that: $1 - y_i < \boldsymbol{x_i}, \boldsymbol{w} > \leq \xi_i; \ \forall i$.

2.2 Multi-labeler Analysis Based on SM-SVM Formulation

To address the matter that we are concerned about in this work, we aim to design a suitable supervised classifier from the information given by different sources (labeling vectors). In this work, we propose to incorporate a penalty factor θ_t, such that $\hat{f}(\cdot)$ decreases when adding right labels otherwise it should increase. This approach is done in a similar way as that proposed in [5] but using a quadratic version. Consider a set of k panelists who assign their corresponding labeling vectors. Then, the t-th panelist is to be associated to penalty factor θ_t, where $t \in [k]$ and $[k] = \{1, \ldots, k\}$. Accordingly, by including the penalty factor $\boldsymbol{\theta}$, we can re-write the functional given in 2 as:

$$\min_{\boldsymbol{w},\boldsymbol{\xi}} \frac{\lambda}{2}||w||_2^2 + \frac{1}{2m}\sum_{i=1}^{m}(\xi_i + \frac{1}{k}\sum_{t=1}^{k} c_{ij}\theta_t)^2,$$

$$\text{s.t.} \quad \xi_i \geq 1 - y_i\langle \boldsymbol{x}_i, \boldsymbol{w}\rangle - \frac{1}{k}\sum_{t=1}^{k} c_{it}\theta_t$$

where c_{it} is the coefficient for the linear combination of all θ_t representing the relevance of the information given by t-th panelist over the sample i. This term decreases the effect of the wrong labels in the construction of the boundary decision. So

$$c_{it} = \frac{\mathrm{n_e}(\boldsymbol{y}^{(t)} = \boldsymbol{y}_{ref})}{m}|\boldsymbol{w}^\top \boldsymbol{x}_i|.$$

Here $\mathrm{n_e}$ is the number of elements that satisfy the condition and Y_ref is the reference labeling vector, normally this parameter is the majority vote of the involved evaluators. In these terms, c_{it} weighted the difference between each evaluator and the reference labeling vector and also, it takes into account the distance of each sample and the decision hyperplane. Defining an auxiliary variable $\hat{\xi}_i$ as

$$\hat{\xi}_i = \xi_i + \frac{1}{k}\sum_{t=1}^{k} c_{it}\theta_t$$

The corresponding $\hat{\xi}_i$ vector formulation is

$$\hat{\boldsymbol{\xi}} = \boldsymbol{\xi} + \frac{1}{k}\boldsymbol{C}\boldsymbol{\theta} \tag{3}$$

Then, the problem formulation will be

$$\min \frac{\lambda}{2}\boldsymbol{w}^\top \boldsymbol{w} + \frac{1}{2m}\hat{\boldsymbol{\xi}}^\top \hat{\boldsymbol{\xi}}, \tag{4}$$

$$\text{s.t.} \quad \hat{\boldsymbol{\xi}} \geq \mathbf{1}_m - (\boldsymbol{X}\boldsymbol{w}) \circ \boldsymbol{y} \tag{5}$$

Assuming the critical case $\hat{\boldsymbol{\xi}} = \mathbf{1}_m - (\boldsymbol{X}\boldsymbol{w}) \circ \boldsymbol{y}$, the Lagrangian of (4) is:

$$\mathcal{L}(\boldsymbol{w}, \hat{\boldsymbol{\xi}}|\lambda) = \frac{\lambda}{2}\boldsymbol{w}\boldsymbol{w}^\top + \frac{1}{2m}\boldsymbol{\xi}^\top \boldsymbol{\xi} + (\boldsymbol{\xi} - \mathbf{1}_m + (\boldsymbol{X}\boldsymbol{w}) \circ \boldsymbol{y})^T \boldsymbol{\alpha} \tag{6}$$

Now, solving the Karush-Kuhn-Tucker conditions, we have:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \lambda \boldsymbol{w} + (\boldsymbol{x}^\top \circ (\boldsymbol{y}^\top \otimes \mathbf{1}_d))\boldsymbol{\alpha} = 0 \Rightarrow \boldsymbol{w} = -\frac{1}{\lambda}(\boldsymbol{x}^\top \circ (\boldsymbol{y}^\top \otimes \mathbf{1}_d))\boldsymbol{\alpha},$$
$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}} = \frac{1}{m}\hat{\boldsymbol{\xi}} + \boldsymbol{\alpha} = 0 \Rightarrow \hat{\boldsymbol{\xi}} = -m\boldsymbol{\alpha}$$
$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}} = \hat{\boldsymbol{\xi}} - \mathbf{1}m + (\boldsymbol{X}\boldsymbol{w}) \circ \boldsymbol{y} = 0,$$

where $\boldsymbol{\alpha}$ is the vector of lagrange multipliers. Under these conditions and eliminating the primal variables from (6), we can pose a new problem in terms of the dual variable $\boldsymbol{\alpha}$, so:

$$\min_{\boldsymbol{\alpha}} \hat{f}(\boldsymbol{\alpha}|\lambda) = \frac{1}{2\lambda}\boldsymbol{\alpha}^\top \boldsymbol{P}\boldsymbol{\alpha} + \boldsymbol{\alpha}^\top \boldsymbol{D}\boldsymbol{\alpha} - \mathbf{1}_m^\top, \quad \text{s.t.} \quad \boldsymbol{\alpha} > 0 \tag{7}$$

with

$$\boldsymbol{P} = (\boldsymbol{x}^\top \circ (\mathbf{1}_d \otimes \boldsymbol{y}^\top))^\top ((\boldsymbol{x}^\top \circ (\mathbf{1}_d \otimes \boldsymbol{y}^\top))),$$
$$\boldsymbol{D} = (-\frac{1}{\lambda}(\boldsymbol{x} \circ (\mathbf{1}_d \otimes \boldsymbol{y}^\top))^\top) \circ ((\mathbf{1}_d \otimes \boldsymbol{y}^\top) \circ \boldsymbol{x}^\top)$$

As it can be appreciated, formulation given by (7) is an evident quadratic problem with linear constraints, which can be solved by means of heuristic quadratic programming methods. Finally, θ value is calculated by this way:

$$\boldsymbol{\theta} = \boldsymbol{C}^\dagger (1 - \boldsymbol{y} \circ (\boldsymbol{X}\boldsymbol{w}) - \boldsymbol{\xi}) \tag{8}$$

As a remarkable matter, it is important to mention that factors $\boldsymbol{\theta}$ is greater as the amount of wrong labels increase, in accordance with equation (8). This implies that the labelers are strongly penalized when their labeling vectors scape from the ratio of the reference label vector, in this case, majority vote vector. Vector $\boldsymbol{\theta}$ also depends on the distance between each sample and the estimated decision boundary. In the case when the labeling vector corresponding to a specific labeler match with the reference vector, the θ value should be 0.

3 Experimental Setup

3.1 Database

For experiments, the Iris database from UCI repository [6] is used. We take the first 100 samples being the two linear separable clases, namely `setosa` and `versicolor` classes. A simulated labeling vector set is built to emulate different labelers. The reference labeling vector is estimated as:

$$\boldsymbol{y}_{\text{ref}} = \text{sign}\left(\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{y}^{(t)}\right),$$

which corresponds to the majority vote.

3.2 Experiments

To test the methodology performance, four experiments are made. The difference between the experiments is the considered error percentage (ε) added to each simulated label set. In the first experiment, the simulated labels are built keeping the same ε for the two classes (balanced error). The second and third experiments use a label simulated set that only include error in a class (unbalanced error), second experiment affects a class and third experiment affects the another class. The last one, uses a combination of error where the second class is affected with the double of error porcentaje that the another class, it is the worst scenario. In all the experiments the ε is varied in the range from 5% to 45% by adding 5% per iteration (that is, a total of nine different ε). Additionally, for each specific ε are simulated 100 labelers. Then, the classifier $\boldsymbol{w}$ and a set of predicted labels are calculated, it is carried out by taking randomly a number t of labeling vectors from the 100 simulated labelers corresponding to this ε. The last procedure is repeated 50 times for each considered error. Predicted labels, allows to make confusion matrixes in relation to simulated labels, for that reason the data mean and standard deviation is calculated for the accuracy, sensibility and specificity from the confusion matrixes, it enable to measure the methodology performance.

4 Results and Discussion

Fig. 1(b) presents the original labels, namely the labels assigned in the Iris database for each class, the Fig. 1(c) give the majority vote, of the iterations with a ε of 20%, in this figure it is evident the description in 3.2 for the first experiment, both classes have the same ε. In Fig. 1(d) the corresponding estimated labels are showed, it is possible to note that the original and estimated labels are similar, it demostrates that the methodology is capable to find, the original labels from a noisy set of expert labels.

Meanwhile. Fig. 1(a) corresponds with the accuracy calculated for the first experiment, the methodology performance maintains a good value until the 35% in εs, it is a good yield because inclusive in high presence of wrong labels, the original labeling vectors are recovered. From a ε superior to 35% the standard deviation grows, but even

it keeps a good result. Values above 40% indicate an intolerable amount of wrong labels and it is a clear indicator of an inexpert evaluator. In Fig. 1(d) also can be seen that two samples have problems to be estimated, this samples correspond with samples nearby to the desicion boundary, just in this samples exists some ambiguity precisely by the nearness between the classes.

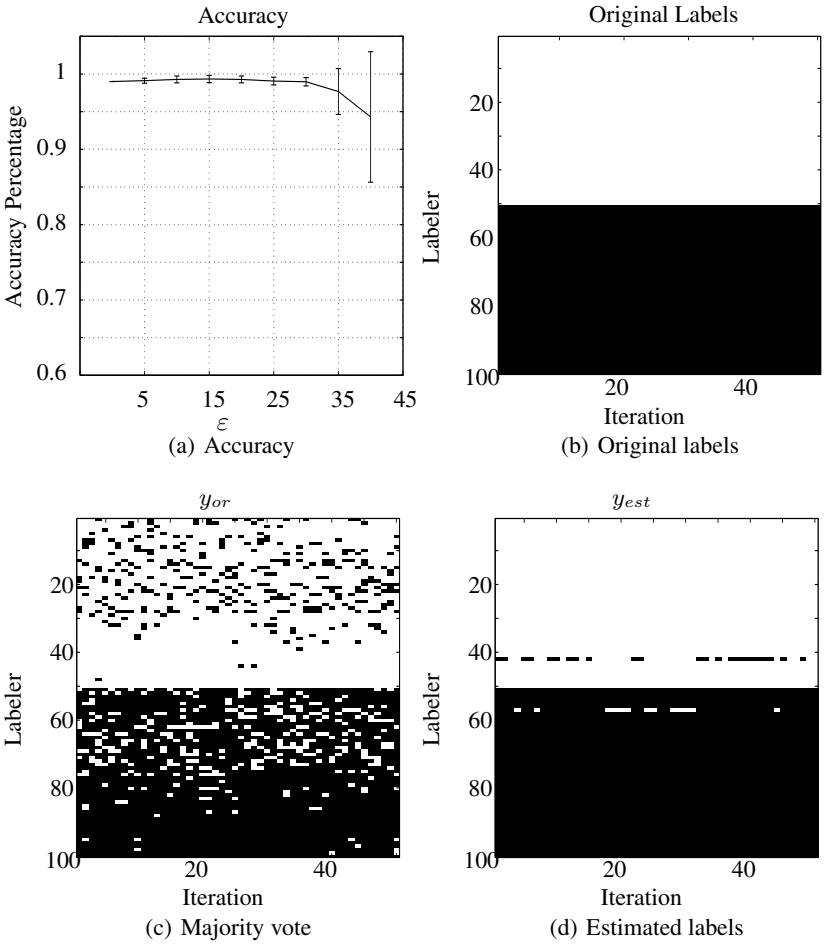


Fig. 1. Experiment1: accuracy; original, majority vote and estimated labels

In Fig. 2 it is important to observe for the experiments 2 and 3, namely, those where the noisy labels are introduced only in a class, that the estimated labels (Figs.2(g) and 2(h)) highly match with the original labels 1(b), it indicates, that when only a class is wrong labeled, it is posible to get a confiable decision boundary, that is also appreciated in table 1 and figures 2(a) and 2(b) where it is possible to find that the accuracy, sensibility and specificity are stables, it is an important result, because in several pattern recognition scenarios for automatic detection of pathologies, is common to find that

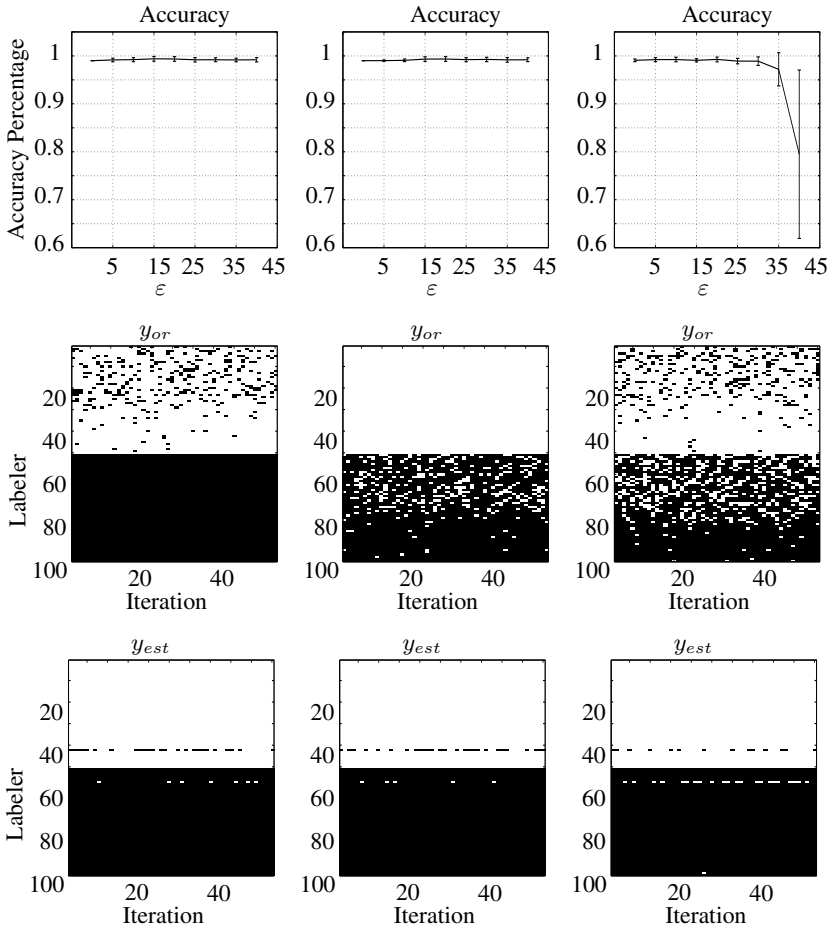


Fig. 2. Experiments Results

labelers (in many times medical experts) can make good diagnosis of normal patients, however the objective class ussually is wrong diagnosed. The present methodology can be used in these environments of automatic detection, because when only a class is wrong labeled, it is posible to build an efficient desicion boundary.

Figures 2(c), 2(f) and 2(i) shows the four experiment, it is the worst scenario for the present methodology, because there are many desbalanced noisy labels, nonetheless, can be apreciated that the estimated labels are correctly calculated inclusive for ε same 30% by a class and superior to 45% for the another class. The accuracy figure shows that only with values up to 40% the dispersion is large enough to consider a bad estimation of the desicion boundary, for this ε the accuracy value is inferior to the 80%, nevertheless, it is important to remember that this is the most difficult training environment for this methodology and with labeling error percentages as large as the exibit in this experiment, it is not possible to work in pattern recognition. Table 1 sumarize the

Table 1. Accuracy, sensibility and specificity experiments 1 and 4

Error percentage	Experiment 1			Experiment 4		
	Accuracy	Sensibility	Specificity	Accuracy	Sensibility	Especificity
0	99.0±0.00	100.0±0.00	98.0±0.00	99.1±0.27	100.0±0.00	98.2±0.55
5	99.1±0.33	100.0±0.00	98.2±0.66	99.2±0.42	100.0±0.00	98.4±0.84
10	99.3±0.45	100.0±0.28	98.6±0.93	99.2±0.48	99.7±0.74	98.8±0.99
15	99.3±0.48	99.5±0.86	99.2±1.00	99.1±0.34	99.6±0.78	98.5±0.89
20	99.3±0.45	99.5±0.89	99.1±1.01	99.3±0.49	99.1±1.01	99.4±0.91
25	99.1±0.51	99.0±1.01	99.2±1.00	**98.9±0.57**	**98.8±1.28**	**99.0±1.01**
30	99.0±0.55	98.7±1.26	99.2±0.98	**98.9±0.89**	**98.4±1.85**	**99.4±0.93**
35	97.7±3.05	96.7±3.95	98.7±2.51	**97.2±3.46**	**96.1±4.64**	**98.3±2.72**
40	94.3±8.66	92.8±9.50	95.8±8.01	**79.5±17.57**	**77.7±17.62**	**81.3±17.92**

Table 2. Accuracy, sensibility and specificity experiments 2 and 3

Error percentage	Experiment 2			Experiment 3		
	Accuracy	Sensibility	Especificity	Accuracy	Sensibility	Especificity
0	**99.0±0.00**	100.0±0.00	98.0±0.00	**99.0±0.00**	100.0±0.00	98.0±0.00
5	**99.2±0.37**	100.0±0.00	98.3±0.74	**99.0±0.14**	100.0±0.00	98.0±0.28
10	**99.2±0.42**	100.0±0.28	98.5±0.86	**99.1±0.27**	100.0±0.28	98.2±0.61
15	**99.4±0.49**	99.8±0.55	99.0±1.01	**99.3±0.48**	99.8±0.61	98.9±1.00
20	**99.4±0.48**	99.7±0.70	99.0±1.01	**99.4±0.48**	99.8±0.61	98.9±1.01
25	**99.2±0.45**	99.4±0.91	99.0±1.01	**99.2±0.42**	99.6±0.84	98.9±1.00
30	**99.2±0.40**	99.1±1.01	99.3±0.96	**99.3±0.44**	99.8±0.66	98.8±0.98
35	**99.2±0.37**	98.8±0.99	99.5±0.86	**99.2±0.48**	99.2±0.99	99.2±1.00
40	**99.2±0.45**	98.8±1.00	99.6±0.84	**99.2±0.40**	99.2±0.98	99.2±1.00

results for experiments 1 and 4, this experiments are similar in the sense of both classes are contaminated with noisy labels. In the two experiments, the desicion boundary estimation is weak for contamination above to 30% where the deviation grows quickly. Specially, the experiment 4 is affected, the bolded values show that when the evaluators have labeling inconsistences for both classes, it is difficult calculate properly the boundaries.

In table 2 the bolded values allow appreciate that for experiments 2 and 3 the accuracy, sensibility and specificity are stable along every ε, this stability is optimal, specially if it is considered that in several pattern recognition problems it is common to find a only one class good labeled. In this table also it is notable the highlighted values, because show that in the conditions of this experiments the sensibility remain in

99% up to 30% of ε, this result is important since the sensibility evaluate as good is the methodology on relation with the objective class, in automatic pathological detection terms, the sensibility measure as the methodology indentify the pathologic patients.

5 Conclusions and Future Work

Experimentally, we proved that the proposed approach is capable to retort the original labels even when the training labels contains a considerable ε level. The penalization term attached to the standard SM-SVM formulation, allows to reduce the effect of the noisy labels in the new formulation training. The experiments presented show that the present methodology is able to be used in aid systems for pathologies detection, principally because is capable of maintain good values for sensibility even when important amount of wrong labels, it favors the detection of instances in the objective class.

It is important to note that, when only a class is affected for the wrong labels the methodology can estimate correctly the original labels, it is remarkable because is normal that evaluators, i.e. doctors, can label correctly a one class, the normal or control class, the pathologic class contains the labeling errors. For the last reason, the present methodology is valid and present an reliable alternative to address multilabeler problems.

For future works, we are aiming to explore alternatives to improve the reference labeling vector setting, since the majority vote may not be an adequate reference for all cases, specially, when there are many supposed wrong labelers. Aditionally, it is necessary review other ways to penalize the error labels oriented to improve the penalty factor in terms of the natural distribution of the data. Finally, the extension of this methodology to the multiclase case is other outstanding contribution to the state of the art.

Acknowledgments. This work is supported by the "Aprendizaje de máquina a partir de múltiples expertos en clasificación multiclase de señales de voz" project associated with "Jóvenes Investigadores" program by COLCIENCIAS and Universidad Nacional de Colombia - Manizales and the project "Servicio de Monitoreo Remoto de Actividad Cardíaca para el Tamizaje Clínico en la red de Telemedicina del Departamento de Caldas" financed with the "Fondo Estampilla Universidad Nacional de Colombia Manizales - Universidad de Caldas".

References

1. Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L.: Learning from crowds. Journal of Machine Learning Research 11, 1297–1322 (2010)
2. Dekel, O., Shamir, O.: Vox populi: Collecting high-quality labels from a crowd. In: Proceedings of the 22nd Annual Conference on Learning Theory (2009)
3. Smyth, P., Fayyad, U.M., Burl, M.C., Perona, P., Baldi, P.: Inferring ground truth from subjective labelling of venus images. In: NIPS 1994, pp. 1085–1092 (1994)
4. Crammer, K., Kearns, M., Wortman, J., Bartlett, P.: Learning from multiple sources. In: Advances in Neural Information Processing Systems, vol. 19 (2007)
5. Dekel, O., Shamir, O.: Good learners for evil teachers. In: ICML, p. 30 (2009)
6. Frank, A., Asuncion, A.: UCI machine learning repository (2010)

New Cues in Low-Frequency of Speech for Automatic Detection of Parkinson's Disease

E.A. Belalcazar-Bolaños[1], J.R. Orozco-Arroyave[1,3], J.F. Vargas-Bonilla[1], J.D. Arias-Londoño[1], C.G. Castellanos-Domínguez[2], and Elmar Nöth[3]

[1] Universidad de Antioquia, Medellín, Colombia
[2] Universidad Nacional de Colombia, Manizales, Colombia
[3] Universität Erlangen-Nürnberg, Erlangen, Germany

Abstract. In this paper, the analysis of low-frequency zone of the speech signals from the five Spanish vowels, by means of the Teager energy operator (TEO) and the modified group delay functions (MGDF) is proposed for the automatic detection of Parkinson's disease.

According to our findings, different implementations of the TEO are suitable for tackling the problem of the automatic detection of Parkinson's disease. Additionally, the application of MGDF for improving the resolution of the speech spectrum in the low frequency zone is able for enhancing differences exhibited between the first two formants from speech of people with Parkinson's disease and healthy controls.

The best results are obtained for vowel /e/, where accuracy rates of up to 92% are achieved. This is an interesting result specially if it is considered that there are muscles that are involved in the production of the vowel /e/ but not in other vowels.

Keywords: Group Delay Functions, Parkinson's disease, Teager Energy Operator.

1 Introduction

The Parkinson's disease (PD) is characterized by the loss of dopaminergic neurons in the mid brain and its main symptoms of PD are tremor, rigidity and other movement disorders. It is demonstrated that about 90% of the people with Parkinson's disease (PPD) also develop speech impairments [1], however only from 3% to 4% of the patients receive speech therapy [2].

Acoustic studies further indicate reduction of the sound pressure level, reduced pitch variability, phonatory instability, increased noise and cycle to cycle variability during phonation of PPD [3]. Movement velocities and displacements of the velum also tend to be reduced in PPD, and acoustic studies suggest increased nasal airflow for speech in about 70% of PPD [4], [5], such increment is characterized for generating alterations in the speech spectrum [6]. Additionally, there are evidences about the alterations in the low frequency zone of the speech spectrum and the excess of nasal airflow during phonations of PPD [6], however, few works have taken advantage of this fact. In [7] the authors divide

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 283–292, 2013.

the speech spectrum into its high and low zones for finding the voice low tone to high tone ration (VLHR) as the quotient between the low frequency power and high frequency power. The VLHR is used for the automatic evaluation of nasalization in speech signals. According to their results the VLHR is a potential quantitative index of hypernasal speech and can be applied in either basic or clinical studies. Likewise, in [8] the authors evaluate the performance of a speech recognition system based on modified group delay functions (MGDF), applied over the low frequency zones in the speech spectrum, for the automatic detection of dysarthria in continuous speech. According to their results, the application of GDF is not suitable for the automatic detection of dysarthric speech signals in continuous speech.

Additionally, the speech impairments in PPD are related to the vocal fold bowing and incomplete vocal fold closure [9], besides nonlinear behavior of vocal fold vibration has already been demonstrated in [10]. Considering these evidences, it is possible to include nonlinear energy operators for characterizing speech from PPD. In [11] the authors use the Teager energy operator (TEO) [12] for calculating signal to noise ratio measures in speech signals from PPD, however such operator has not been directly used as a measure of the speech for the automatic detection of PD.

With the aim of include the analysis of the low frequency zone of the speech spectrum and the nonlinear behavior of the vocal folds vibration, in this work the use of MGDF and TEO is proposed as a new method for the automatic classification of speech signals from PPD and HC.

The rest of the paper is organized as follows. In section 2, a methodology proposed is presented, section 3 provides details about the experimental framework with some experimental details about the followed methodology for classification and error estimation, section 4 show the graphics and tables with success classification rates. Finally, in section 5, some conclusions are provided.

2 Metodology

Figure 1 shows a schematic of the methodology used in this work. Every stage of the process will be explained in the following.

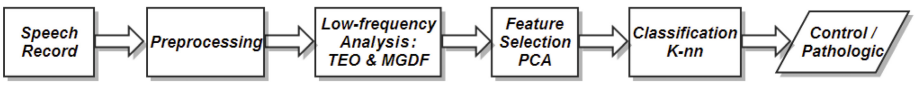


Fig. 1. Methodology

Preprocessing and low frequency analysis. Speech recordings are preprocessed by means of a short temporal analysis using windows of 40ms length with an overlap of 20ms. After, each frame is characterized using a low frequency analysis that is performed through the implementation of TEO and MGDF.

The TEO is a nonlinear operator that is used for taking advantage of the existence of multiple components on a signal $x(n)$ and it is given by [12]:

$$\Psi[x(n)] = x(n)^2 - x(n+1)x(n-1) \tag{1}$$

In the case of nasalized voice signals it is demonstrated that the speech spectrum has multiple components due to the presence of nasal formants and anti formants [6]. The voice signal $x(n)$ can be expressed as the sum of two uncorrelated components $s(n)$ and $g(n)$, when the definition 1 is applied over a multi-component signal $x(n) = s(n) + g(n)$ it is obtained the following expression [13]:

$$\Psi[x(n)] = \Psi[s(n)] + \Psi[g(n)] + \Psi_{cross}[s(n), g(n)]\Psi_{cross}[g(n), s(n)]$$

where $\Psi_{cross}[s(n), g(n)] = g(n)s(n) - g(n+1)s(n-1)$. Note that TEO does not obey the superposition principle because additional terms appear when it is applied over a multi-component signal. Considering this fact, it is possible to apply different measurements that reflect differences in the estimation of the TEO for multi-component signals and for single-component signals.

There are other two different ways for the estimation of the TEO, one is based on the fast Fourier transform (FFT) and defines the TEO as follows [14]:

$$\gamma_n = \left[\sum_{i=1}^{\Omega} i^2 S_n(i)\right]^{\frac{1}{2}}$$

where $S_n(i)$ is the power spectral density from the n speech frame, calculated using the FFT, and i is the frequency value in the discrete domain.

Another estimation is presented in [15] as the generalized version of the TEO and the expression is as follows:

$$\Psi[x(n)] = x(n)^{\frac{2}{m}} - [x(n-M)x(n+M)]^{\frac{1}{m}} \tag{2}$$

Where M and $m \in \mathbb{Z}$. Note that the expression 2 allows the variation of exponents (m) and delays (M).

The different versions of the TEO exposed above can be used for taking advantage of the existence of additional formants in the voice signals from PPD due to the excess of nasalization [6]. One healthy voice only has vocal formants, thus its spectrum can be expressed as $S_{HC} = \sum F(\omega)$, where $F(\omega)$ represents the oral formants. A speech recording from PPD, with excess of nasalization, has oral formants, anti-formants and nasal formants, thus its spectrum can be represented by $S_{PPD} = \sum F(\omega) - \sum AF(\omega) + \sum NF(\omega)$, where $AF(\omega)$ represents the anti-formants and $NF(\omega)$ the nasal formants. When a healthy voice is filtered by means of a low pass filter, it is possible to extract its first formant F_1, however, if a speech signal with excess of nasalization is filtered, the result will contain F_1 along with two additional components due to the existence of anti-formants and nasal formants. If instead a bandpass filter is used (with cutoff bands around F_1), both the result for healthy and nasalized voices will contain only F_1. The differences between both bandpass and lowpass filtered signals can

be exploited for the automatic detection of excess nasalization in speech from PPD by means of the implementation of different measures oriented to detect such differences.

The first measurement is Correlation Teager Energy Operator ($CTEO$) [13]. It is calculated as the correlation between the TEO estimated for the low-passed and band-passed filtered signals, the result of this measure is expected to be high for the case of healthy voices (both filtered signals contained only F_1), but if the correlation is calculated for pathological voices, the result is expected to be low due to the existence of additional nasal formants and anti-formants in the low passed filtered signal.

The other three measurements are the Euclidean distance (ED), the logarithmic distance (LD) and the area under the TEO estimates (IA). All of these measurements are calculated between the TEO estimated for the low-passed and band-passed filtered signals. In all cases, it is expected to obtain higher values of the measures in the case of nasalized voices due to the existence of additional components that introduce differences between both estimations of the TEO.

On the other hand, the MGDF are applied to improve the resolution of the low frequency zone in the speech spectrum. The group delay function is defined as $\tau(\omega) = -\frac{\partial\theta(\omega)}{\partial\omega}$, and for discrete time signals it can be computed as [16]:

$$\tau(K) = \frac{X_R.Y_R + X_I.Y_I}{|X(K)|^2} \tag{3}$$

where $X(K)$ and $Y(K)$ are the N-point discrete Fourier transform (DFT) of the sequences $x(n)$ and $nx(n)$, respectively. The subscripts R and I denote the real and imaginary parts, respectively.

To reduce the spiky nature of the group delay function, which is due to the pitch peaks, noise, and windowing effects , the original function is modified as [17].

$$\tau(K) = \mathbf{sign}\left|\frac{X_R.Y_R + X_I.Y_I}{S(K)^{2\gamma}}\right|^{\alpha}$$

where $S(K)^2$ is a cepstrally smoothed version of $|X(K)|^2$ and the **sign** is the original sign of the group delay function given in 3. To derive the a smoothed group delay spectrum, the values of α and γ should be less than 1 [17].

In the process for the improvement of the signal's spectrum resolution the voice signal is low-pass filtered with a cutoff frequency of $800Hz$, then the MGDF is calculated for the filtered spectrum and finally the first two formants are estimated. The first formant (F_1) is found below the $500Hz$ and the second formant (F_2) is between $500HZ$ and $800Hz$. For the case of speech signals with excess of nasalization, the amplitude of F_1 will be lower than in the case of healthy voices, thus the ratio between the amplitudes of F_1 and F_2, called $\tau(F_1)$ and $\tau(F_2)$, can be used as an index of the presence of nasalization in the speech signal [17]. Such index is called group delay function-based acoustic measure ($GDAM$) and is defined as:

$$GDAM = |\tau(F_1)| \,/\, |\tau(F_2)|$$

Automatic Features Selection and Classification. The selection of features is addressed through the application of principal components analysis (PCA). It is a statistical technique applied here to find out a low-dimensional representation of the original feature space, searching for directions with greater variance to project the data. Although, PCA is commonly used as a feature extraction method, it can be useful to properly select a relevant subset of original features that better represent the studied process [18]. In this sense, given a set of features $(\boldsymbol{\xi}_k : k = 1, \ldots, p)$ corresponding to each column of the input data matrix $\mathbf{X}$, the relevance of each $\boldsymbol{\xi}_k$ can be analyzed for finding the resulting subspace $\mathbf{Y}$. More precisely, relevance of $\boldsymbol{\xi}_k$ can be identified looking at $\boldsymbol{\rho} = \begin{bmatrix} \rho_1 \ \rho_2 \cdots \ \rho_p \end{bmatrix}^\top$, where $\boldsymbol{\rho}$ is defined as $\boldsymbol{\rho} = \sum_{j=1}^{m} |\lambda_j \mathbf{v}_j|$. ($\lambda_j$ and $\mathbf{v}_j$ are the eigenvalues and eigenvectors of the initial matrix, respectively). Therefore, the main assumption is that the largest values of ρ_k point out to the best input attributes, since they exhibit higher overall correlations with principal components.

The decision of whether a voice recording is from PPD or HC is taken with a K nearest neighbor (K-nn) classifier. Considering that the aim of this work is to analyze the discrimination capability of the described features, this classifier is chosen because of its simplicity allowing us to focuses our analysis to the features and not to the classifier.

3 Experimental Framework

Database. The data for this study consists of speech recordings from 20 PPD and 20 HC sampled at $44.100Hz$ with 16 quantization bits. All of the recordings were captured in a sound proof booth. The people that participated in the recording sessions are balanced by gender and age: the ages of the men patients ranged from 56 to 70 (mean 62.9 $\pm$ 6.39) and the ages of the women patients ranged from 57 to 75 (mean 64.6 $\pm$ 5.62). For the case of the healthy people, the ages of men ranged from 51 to 68 (mean 62.6 $\pm$ 5.48) and the ages of the women ranged from 57 to 75 (mean 64.8 $\pm$ 5.65). All of the PPD have been diagnosed by neurologist experts and none of the people in the HC group has history of symptoms related to Parkinson's disease or any other kind of movement disorder syndrome.

The recordings consist of sustained utterances of the five Spanish vowels, every person repeated three times the five vowels, thus in total the database is composed of 60 recordings per vowel on each class. This database is built by *Universidad de Antioquia* in Medellín, Colombia.

Experimental Setup. The voice recordings were segmented and windowed using frames of $40ms$ with an overlap of $20ms$. The characterization of speech recordings is made considering two versions of the TEO, one is the generalized form (TEO_1) as indicated in the equation 2 and the other one is such based on the FFT (TEO_2). In the case of TEO_1, different values of M and N are tested. In the filtering process (band-pass: BW_1 and low-pass: BW_2), different values of

bandwidth were also tested. In table 1 the results after an exhaustive search of the optimal values for M, N, BW_1 and BW_2, are indicated for each Spanish vowel. The same optimal values found for the bandwidths of the low-pass and band-pass filters are also used in the estimation of the measurements based on the TEO_2.

Table 1. Optimal values of bandwidths, exponent m and delay M for the estimation of TEO_1

Vowel	m	M	BW_1	BW_2
/a/	4	3	25	450
/e/	5	5	25	450
/i/	2	3	300	550
/o/	5	5	25	450
/u/	2	5	300	550

The full set of features is formed by 13 measures: the four measurements taken from both estimations of the TEO, the frequency values of the first and second formants (F_1 and F_2) calculated using the MGDF, the amplitude values of the first and the second formants ($\tau(F_1)$ and $\tau(F_2)$) also calculated using the MGDF and the $GDAM$ index. Each measure is obtained for every frame in the voice signal and after, four statistics are estimated per measure (mean value, standard deviation, kurtosis and Sweness), forming a total of 52 measures for representing each voice recording. Table 2 summarizes the set of features considered in this work and the indexes assigned for each one.

Table 2. Index allocation for features

	IA		$CTEO$		ED		LD		$\tau(F_1)$	$\tau(F_2)$	F_1	F_2	$GDAM$
	TEO_1	TEO_2	TEO_1	TEO_2	TEO_1	TEO_2	TEO_1	TEO_2					
Mean	1	2	3	4	5	6	7	8	9	10	11	12	13
Std	14	15	16	17	18	19	20	21	22	23	24	25	26
Kurtosis	27	28	29	30	31	32	33	34	35	36	37	38	39
Skewness	40	41	42	43	44	45	46	47	48	49	50	51	52

* IA :Area under the TEO curves, ED : Euclidean Distance, $CTEO$: Correlation Teager energy operator LD : Logarithmic Distance, $GDAM$: Group Delay Acoustic Measure

The tests performed over the proposed system have been made following the strategy indicated in [19]. The 70% of the data are used for the feature selection and for training the classifier and the remaining 30% is for testing; ten different subsets for training and testing are randomly formed in ten repetitions of the randomization of the data, in order to perform a total of ten independent experiments, each one with its results, allowing the calculation of confidence intervals for the general performance analysis and robustness of the proposed system.

4 Results and Discussion

The automatic features selection process based on PCA gives the resulting sub set of features that better represent the phenomena and also gives the order

Table 3. Indexes of selected features after PCA transformation and correlation analysis

Vowels	Feature Index															
/a/	32	12	3	20	15	26	11	46	13	2	51	29	31	1	42	14
	10	40	49	48	38	50	36	52	9	35	22	23	37			
/e/	2	31	43	12	7	15	16	27	13	24	44	14	11	1	49	37
	50	8	21	51	39	10	9	22	48	35	23					
/i/	19	2	5	40	11	31	42	12	10	39	51	26	14	23	50	49
	9	37	22	48	46	35	33	20	7							
/o/	17	5	20	30	11	25	13	33	14	27	1	42	26	40	36	35
	23	38	37	48	51	52	9	50	49	10	22					
/u/	5	34	40	42	22	14	12	18	27	25	26	52	29	49	37	11
	31	13	9	23	36	10	51	50	48	46	35	33	20	7		

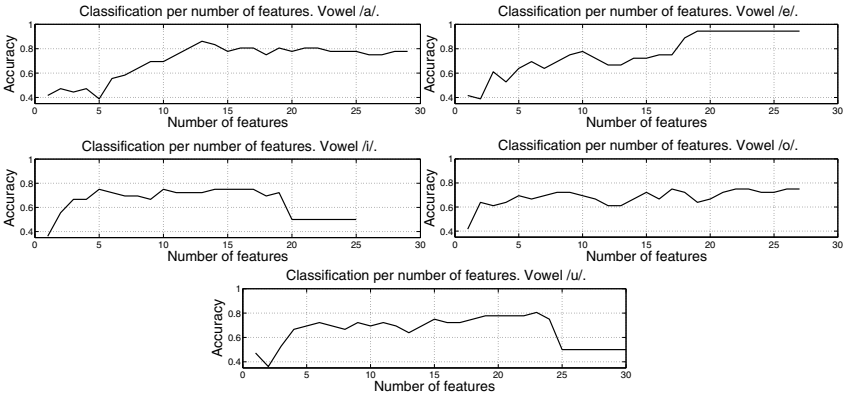


Fig. 2. Success rates obtained per vowel

of features according to their contribution in terms of the cumulative variance. Table 3 indicates which of the features remain after the features selection process per vowel. Considering that the dimensionality of the original representation space is 52, the reductions achieved with selection process ranged from 40% to 54%. Figure 2 shows the accuracy results obtained per vowel when each of the selected feature is progressively added to the classification process. The order in which each feature is added is presented in table 3.

Note that the results obtained with the vowel /e/ are better than those obtained with the other vowels. This is an interesting result specially if it is considered that the risorius muscles are involved in the phonation of vowels /e/ and /i/, but not in other vowels [20].

For the case of vowels /i/ and /u/ the accuracy rates fall when more than 20 and 25 features are considered. The results with the vowels /a/ and /o/ are more stable, with accuracies around 80%. With the aim of indicating the results in a more compact way, in table 4 the results obtained per vowel are presented in

terms of accuracy, specificity and sensitivity. Note that the performance achieved with the vowel /e/ is around 12 percentage points better than those obtained with the other vowels and in such case. This result can also be noted in the receiver operating characteristic (ROC) curves that are shown in figure 3. These kind of curves are widely used in clinical applications and the area under such curves (AUC) is considered as a good statistic for representing the general performance of the system [19]. The AUC obtained per vowel are: AUC /a/: 0.7872, **AUC /e/: 0.9214**, AUC /i/: 0.7954, AUC /o/: 0.7799 and AUC /u/: 0.8111.

Table 4. Performance measures

Vowel	Accuracy	Specificity	Sensitivity
/a/	0.7667±0.0631	0.7905 ±0.0821	0.7523±0.0639
/e/	**0.9250±0.0541**	**0.9092±0.0852**	**0.9559±0.0474**
/i/	0.7778±0.0434	0.7648±0.0516	0.8056±0.0715
/o/	0.7778±0.0393	0.8079±0.0672	0.7616±0.0517
/u/	0.7972±0.0586	0.7732±0.0592	0.8293±0.0640

* The results are presented in terms of mean value ± standard deviation.

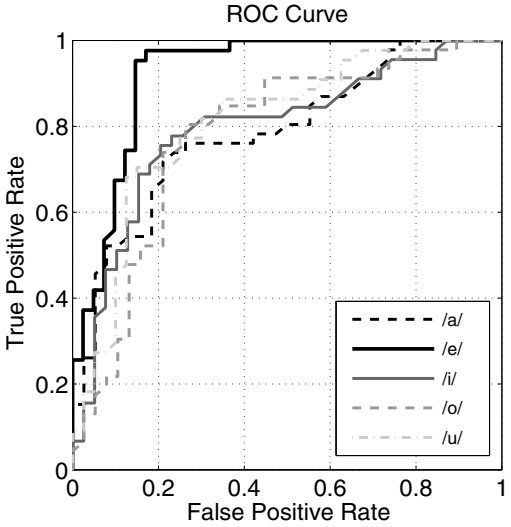


Fig. 3. ROC curves with the best results obtained per vowel

5 Conclusions

A new methodology for the automatic detection of Parkinson's disease, based on the analysis of the low frequency zone of the speech spectrum of the fiver Spanish vowels, is presented.

Different implementations of the Teager Energy Operator are suitable for tackling the problem of the automatic detection of PD. Additionally, the application of Group Delay Functions for improving the resolution of the speech spectrum in the low frequency zone is able for enhancing differences exhibited between the first two formants from speech of PPD and HC.

The stage of automatic features selection allowed the reduction of the dimensionality of the spaces in up to 52%, indicating that there were a lot of redundant information in the original representation space.

According to the obtained results, it is possible to achieve accuracy rates of up to 92% when the vowel /e/ is evaluated. For the other vowels the accuracy results are near to 80%. This difference suggest a more detailed analysis of the set of muscles and/or tissues such as the risorious one, which are involved for the production of the vowel /e/.

Acknowledgments. Juan Rafael Orozco Arroyave is under grants of "Convocatoria 528 para estudios de doctorado en Colombia, generación del bicentenario, 2011" financed by COLCIENCIAS. The authors give a special thanks to all of the patients and collaborators in the foundation "Fundalianza Parkinson-Colombia". Without their valuable support it would be impossible to address this research.

References

1. Ho, A., Iansek, R., Marigliani, C., Bradshaw, J., Gates, S.: Speech impairment in a large sample of patients with parkinson's disease. Behavioral Neurology 11, 131–137 (1998)
2. Ramig, L., Fox, C., Shimon, S.: Speech treatment for parkinson's disease. Expert Review Neurotherapeutics 8(2), 297–309 (2008)
3. McNeil, M.: Clinical Management of Sensorimotor Speech Disorders, 2nd edn. Thieme, New York (2009)
4. Tjaden, K.: Speech and swallowing in parkinson's disease. Top Geriatr Rehabilitation 24(2), 115–126 (2008)
5. Aronson, A., Bless, D.: Clinical voice disorders, 4th edn. Thieme, New York (2009)
6. Kent, R., Weismer, G., Kent, J., Vorperian, H., Duffy, J.: Acoustic studies of dysarthric speech: methods, progress, and potential. Journal of Communication Disorders 32(3), 141–180 (1999)
7. Lee, G., Wang, C., Yang, C., Kuo, B.: Voice low tone to high tone ratio: A potential quantitative index for vowel [a:] and its nasalization 53(7), 1437–1439 (2006)
8. Vijayalakshmi, P., Reddy, M.: Assessment of dysarthric speech and an analysis on velopharyngeal incompetence. In: Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 3759–3762 (2006)
9. Perez, K., Ramig, L., Smith, M., Dromery, C.: The parkinson larynx: tremor and videostroboscopic findings. Journal of Voice 10(4), 353–361 (1996)

10. Giovanni, A., Ouaknine, M., Guelfucci, R., Yu, T., Zanaret, M., Triglia, J.: Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds. Journal of Voice 13(4), 456–476 (1999)
11. Tsnas, A., Little, M., McSharry, P., Spielman, J., Ramig, L.: Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease 59(5), 1264–1271 (2012)
12. Kaiser, J.F.: On a simple algorithm to calculate the "energy" of a signal. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 381–384 (1990)
13. Cairns, D., Hansen, J., Riski, J.: A noninvasive technique for detecting hypernasal speech using a nonlinear operator. IEEE Transactions on Biomedical Engineering 43(1), 35 (1996)
14. Ying, G., Mitchell, C., Jamieson, L.: Endpoint detection of isolated utterances based on a modified teager energy measurement. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 732–735 (1993)
15. Eivind, K.: Signal processing using the teager energy operator and other nonlinear operators (2003)
16. Yegnanaryana, B., Duncan, G., Murthy, H.: Formant extraction from group delay function. IEE Colloquium on Speech Processing, 2/1 –2/4 (1988)
17. Vijayalakshmi, P., Reddy, M., O'Shaughnessy, D.: Acoustic analysis and detection of hypernasality using a group delay function. Transactions on Biomedical Engineering 54(4), 621–629 (2007)
18. Daza-Santacoloma, G., Arias-Londoño, J., Godino-Llorente, J., Sáenz-Lechón, N., Osma-Ruiz, V., Castellanos-Domínguez, C.G.: Dynamic feature extraction: an application to voice pathology detection. Intelligent Automation and Soft Computing 15(4), 665–680 (2009)
19. Sáenz-Lechón, N., Godino-Llorente, J., Osma-Ruiz, V., Gómez-Vilda, P.: Methodological issues in the development of automatic systems for voice pathology detection. Biomedical Signal Processing and Control 1, 120–128 (2006)
20. Phonetics, D.: Dissection of the speech production mechanism. Working Papers in Phonetics, UCLA (102), 1–89 (2002)

Computer-Aided Detection of Microcalcifications in Digital Mammograms to Support Early Diagnosis of Breast Cancer

Nayid Triana and Alexander Cerquera

Faculty of Electronic and Biomedical Engineering, Complex Systems Research Group, Antonio Nariño University, Bogota, Colombia
{natriana,alexander.cerquera}@uan.edu.co

Abstract. Early detection of microcalcifications in mammograms is considered one of the best tools to prevent breast cancer. Although traditionally this task have been performed with analog mammograms, digital mammograms are currently an alternative for examination of breast to detect microcalcifications and any other kind of breast abnormalities. Digital mammography presents some advantages in comparison to its analog counterpart, such as lower radiation dosage for acquisition and possibility to storage for telemedicine purposes. Nevertheless, digitalization entails loss of resolution and difficulties to detect microcalcifications. Therefore, several methods based on digital image processing have been proposed to perform detection of microcalcifications in digital mammograms, to support the early detection and prognosis of breast cancer. However, sometimes computer-aided methods fail due to the characteristics of certain microcalcifications that are hard to detect either by visual examination and by computerized analysis. For this reason, this work presents a method based on contrast enhancement and wavelet reconstruction oriented to increase the rate of computer-aided detected microcalcifications. The images correspond to the mini-MIAS database, which provides mammograms of healthy women and with breast microcalcifications, including the respective coordinates of their locations. The work includes also the application of the method in resolution-enhanced mammograms via sparse representation, with the aim to determine the role of resolution enhancement for a possible improvement in the performance of the method.

1 Introduction

Breast cancer is one of the most principal causes of death in women around the world, whose diagnosis in its early stages increases the opportunities for a patient to enhance her recovery expectations. One of the main ways for its early diagnosis is the detection of microcalcifications in mammograms by visual examination, which appear in form of clusters with diameter between 0.2 to 0.3 mm. Relevant microcalcification clusters are formed by more than four microcalcification spots inside a volume of 1 cm^3; a less number rarely entails to detection of breast cancer [1].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 293–302, 2013.

Analog mammograms are the most employed source of information for detection of microcalcifications, which are basically X-ray images of both breast. However, digital mammograms are also widely utilized by specialists because have certain advantages over analog mammograms, especially the low dose of radiation used for its acquisition and the possibility to store huge quantities of images in digital media for purposes in telemedicine. Nevertheless, in both cases their visual examination depends on the experience and skills of the specialists that observes the images [2]-[4], getting more difficult in digital mammograms because their low contrast and resolution in comparison with their analogue counterpart.

Different works have been presented during the last two decades applying techniques based on digital image processing to develop computer-aided detection systems of microcalcifications. A wide number of methods available in the literature have been applied for this purpose. These methods include contrast enhancement by filtering for enhancement of microcalcifications [5], statistical characterization of microcalcifications and their background tissues [6], characterization of spatial intensity distributions [7], wavelet transform [8], genetic algorithms [9], binary classification by machine learning [10], among others.

Most of these works have shown a promising performance in the detection of microcalcification clusters. However, some of them may tend to overemphasize the sensitivity in the detection at expense of specificity, resulting in an increase of false positive rates [7]. In addition, heterogeneous characteristics of breast tissue among women have influence in the detection of microcalcifications in mammograms acquired to different patients. Likewise, loss of resolution caused by digitalization of analog mammograms can affect the performance of these methods.

In this work, a method based on contrast enhancement and wavelet analysis is proposed to detect microcalcification clusters in the images of the mini-MIAS database, specially in mammograms with presence of microcalcifications hard to find by visual examination. Likewise, it was determined the influence of resolution enhancement of mammograms in the performance of the method, as well as of the type of wavelet utilized in the analysis. The results were evaluated based on calculations of sensitivity and specificity in both non-enhanced resolution and enhanced-resolution images.

2 Materials and Methods

2.1 Database

The digital mammograms analyzed correspond to the mini-MIAS database (Mammographic Image Analysis Society) constructed by the UK National Breast Screening Programme [11]. This database contains 297 images acquired to healthy women and 25 images of patients with microcalcification clusters; mammograms with other kind of abnormality were not included in this work. The documentation of the database indicates the sizes and coordinates of the

microcalcification clusters located previously by specialists. Each image has a resolution of 200 micron pixel edge and a size of 1024 x 1024 pixels.

2.2 Preprocessing

Segmentation. This task was performed to select only the breast area, with the aim to reduce the preprocessing time. Depending on the characteristics of breast tissue in each image, this was done either manually or automatically. In the manual way the function `roipoly` of Matlab® was utilized constructing a mask with polygonal form around the breast area. By the automatic way, the histogram was equalized with a subsequent binarization of the image using automatic thresholding. Afterward, the non-zero pixels were labeled with the function `bwlabel` to segment the breast with `regionprops`.

Contrast Enhancement. The first step in the contrast enhancement was to apply unsharp masking with masks of 3 x 3, emphasizing the high frequency features of the images that allowed to improve the borders and details information. The evaluation of the contrast enhancement was performed by visual inspection on each image. The masks employed are shown in Figure 1:

0	-1	0
-1	4	-1
0	-1	0

(a)

-1	-1	-1
-1	8	-1
-1	-1	-1

(b)

1	-2	1
-2	4	-2
1	-2	1

(c)

Fig. 1. Masks employed for unsharp masking

In most of the images it was necessary to employ the mask of Figure 1(b). The application of unsharp masking was based on:

$$y(n,m) = x(n,m) + \lambda z(n,m) \tag{1}$$

where $z(n,m)$ is the output image from a high-pass linear filter, $x(n,m)$ is the image input, λ is a factor that controls the enhancement level of contrast, and $y(n,m)$ is the output image with enhanced contrast.

The unsharp masking was complemented with a histogram modification, allowing to obtain an image with more emphasized microcalcification in relation to the rest of the image. This task required a threshold to separate the intensity pixels corresponding to microcalcifications and those belonging to the surrounding tissue [14]. Two mathematical functions were employed: linear modification (Equation 2) and exponential modification (Equation 3).

$$P(i,j) = \begin{cases} S(i,j), & si\ S(i,j) \geq T \\ S(i,j) - \bar{D}, & si\ S(i,j) < T \end{cases} \tag{2}$$

$$P\left(i,j\right)=\begin{cases} e^{S(i,j)/k}, & si\ S\left(i,j\right)\geq T \\ e^{S(i,j)/k}-\bar{D}, & si\ S\left(i,j\right)<T \end{cases} \tag{3}$$

where $P(i,j)$ is the output pixel, $S(i,j)$ is the input pixel, T is the threshold, D is a positive integer number, and the k-factor is given by Equation 4.

$$k=\frac{S\left(i,j\right)_{max}}{\ln S\left(i,j\right)_{max}} \tag{4}$$

The values of T and D were adjusted by trial and error according to the observation of the images as these values were modified. Finally, the values of T=210 and D=50 were established.

Resolution Enhancement. The method for resolution enhancement in this work was Image Super-Resolution via Sparse Representation [12]. The resolution of the images was not increased more than two times because the high computational cost of this algorithm. This technique suggests that image patches can be represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary. So, the algorithm provides a sparse representation for each patch of the low-resolution mammogram, and then uses the coefficients of this representation to generate the high-resolution image. All tasks of preprocessing described above (segmentation and contrast enhancement) were also performed over the enhanced-resolution mammograms with this technique.

2.3 Detection of Microcalcifications

The detection of microcalcifications in both, non-enhanced- and enhanced-resolution images, was performed by multiresolution decomposition with discrete wavelet transform of two dimensions. Different combinations of wavelet were tested in this work: Haar, Daubechies, Symlets, Meyer, Coiflets, Biorthogonal and Reverse Biorthogonal. The multiresolution decomposition in each image allowed to obtain four images for a decomposition level: an approximation subband with low-frequency information of the image and three details subbands with high-frequency images. All approximation coefficients were suppressed making them equal to zero [15].

The following step was to filter the details information on each sub-image [16], emphasizing the mammographic characteristics whereas the noise was reduced and allowing to improve those regions of the image with low contrast. Subsequently, a nonlinear filter was applied to improve the high-frequency information in the different levels of decomposition. The subbands filter $FS(i,j)$ is represented in Equation 5:

$$FS\left(i,j\right)=\begin{cases} f\left(i,j\right)-(k-1)T, & f\left(i,j\right)<-T \\ kf\left(i,j\right), & f\left(i,j\right)\leq T \\ f\left(i,j\right)+(k-1)T, & f\left(i,j\right)>T \end{cases} \tag{5}$$

where $f(i, j)$ is the details function of orientation (horizontal, vertical or diagonal) in a decomposition level, k is the gain and T is the threshold. The parameters of gain and threshold were likewise selected by trial and error method as the resultant images were obtained. These were established in k=20 and T=0.5 in each decomposition level. After subbands filtering, it was calculated the discrete wavelet transform in two dimensions with emphasized information in high-frequency. Subsequently, the image was thresholded to remove information that did not correspond to microcalcifications by means of global thresholding:

$$B(x, y) = \begin{cases} 1, si\ C(j, k) \geq U \\ 0, si\ C(j, k) < U \end{cases} \tag{6}$$

where U is a constant that allows either to reduce or increase the threshold. The method was applied in every non-enhanced- and enhanced-resolution image. Afterward, the procedure for detection of the microcalcification clusters was to scan from left to right and from up to down the image in areas of 1 cm^2 per iteration, advancing steps of a pixel in each of them. At the end of the search, it was detected the area with the largest number of microcalcifications. If the detected area contained more than four microcalcification spots, it was defined as the location of the microcalcification clusters.

Testing of the Method. The rates of true positives, false positives, true negatives, and false negatives, were calculated with the aim to determine information of specificity and sensitivity for construction of ROC diagrams. *Positives* were defined as those images with microcalcifications according to the documentation of the database, whereas *negatives* the images of healthy subjects.

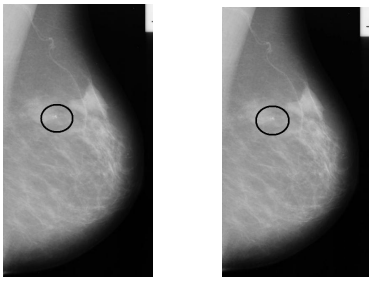
3 Results

Figure 2 shows the result of the segmentation performed in the image named *mdb248* in the database. This procedure allowed to obtain the breast region area in each image, focusing the microcalcifications search only over this region. Figure 3 shows the effect of the resolution enhancement procedure with the same image presented in Figure 2, indicating the presence of the microcalcification cluster inside the black circle. In Figure 4 it can be observed the result of the procedure for contrast enhancement based on unsharp masking and histogram modification, where the microcalcifications got clearly emphasized in the left superior quadrant of the image. During the visual examination in each image, the best results were obtained with exponential modification.

In Figure 5 the squares represent the area where the method detected the microcalcification cluster in the image. It can be observed its location very close to the coordinates provided by the documentation indicated by the circles. Figure 6 shows the result of the procedure in a mammogram with microcalcifications hard to detect by visual examination, where the automated location is relatively close to the labeled location according to the documentation.

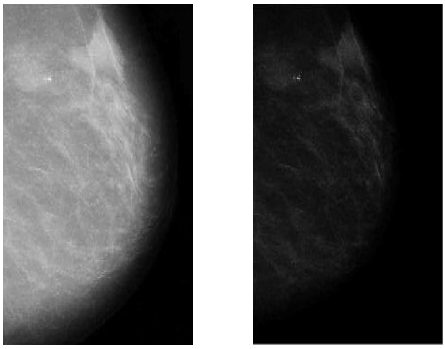
(a) (b) (c)

Fig. 2. a) Original image; b) Segmentation of the breast area; c) Breast area in the mammogram with well defined contours.



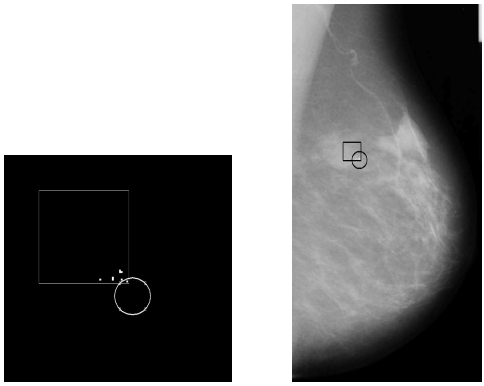
(a) (b)

Fig. 3. a) Image with original resolution (1024x1024 pixels). b) The same image with enhanced resolution via Sparse Representation (2048x2048 pixels). The cluster of microcalcifications are fenced with the black circle.



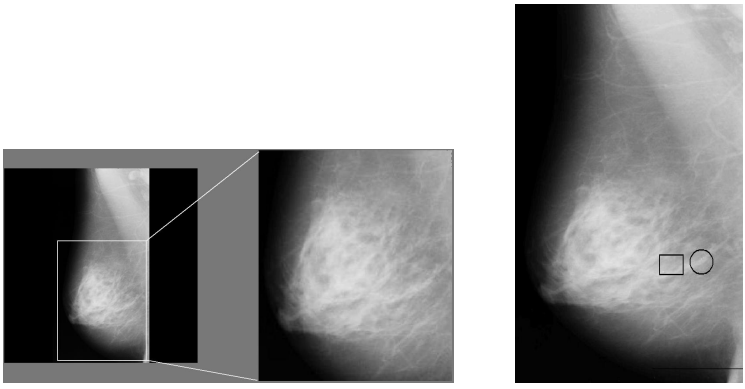
(a) (b)

Fig. 4. a) Image obtained after unsharp masking. b) Image after histogram modification.



(a) (b)

Fig. 5. a) Detected microcalcification cluster with a wavelet *Biorthogonal 3.3*. The circle indicates the coordinates and size provided by the documentation of the database, whereas the square indicates the area of 1 cm^2 where they were automatically detected. b) Superposition of the detected cluster of microcalcifications in the original image.



(a) (b)

Fig. 6. a) Mammogram with microcalcifications hard to detect by visual examination, named as *mdb211* in the database. The region with the microcalcification is augmented. b) Location of the detected microcalcification with a wavelet *Biorthogonal 3.3*. As shown in Figure 5, the circle indicates the coordinates according to the documentation and the square the location provided by the method.

In almost all the wavelet analysis done in this work, it was observed that the performance of the method was better with non-enhanced resolution images than with enhanced-resolution images, such as shown in the ROC curves of Figure 7. Due to space limitations, only the results with the wavelet *Biorthogonal 3.3* of four decomposition levels are shown for this case. Other wavelets were also

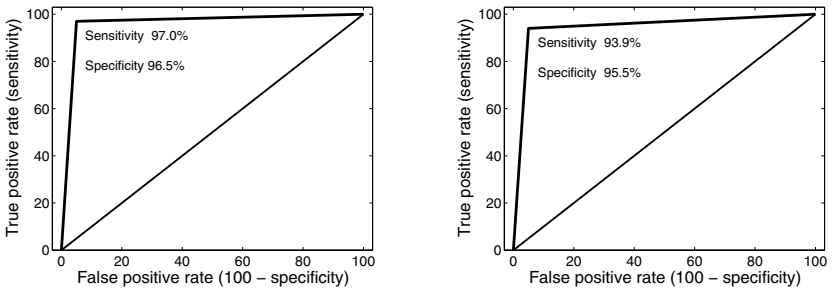


Fig. 7. ROC curves by applying the wavelet *Biorthogonal 3.3*, indicating the percentages of sensitivity and specificity in each case. a) non-enhanced resolution images. b) enhanced resolution images.

applied, being their performances slightly lower than with the *Biorthogonal 3.3*. The exception to this situation was observed applying the wavelet *Daubechies 20*, where the detection of microcalcifications with enhanced-resolution images was better than with non-enhanced resolution images. Nevertheless, its performance was diminished in comparison to the analysis with *Biorthogonal 3.3* (Figure 8). Table 1 shows the number of mammograms detected according to their class (with and without microcalcifications), resolution (non-enhanced and enhanced) and wavelet wave employed. The number of mammograms with microcalcifications in the database is 25, whereas the number without microcalcifications is 297.

4 Discussion

This work presents a method for computer-automated detection of microcalcification clusters tested with the mini-MIAS database. Due to the different characteristics of the breast tissue density among images, the preprocessing effects were assessed by visual inspection in each image, with the aim to match the more adequate procedure for contrast improvement in each case. Therefore, it is difficult to suggest a generalized method that detects a mammogram with microcalcifications in a complete automatic way for all images of the database, especially in the tasks of contrast enhancement.

The analysis with the wavelet *Biorthogonal 3.3* allowed to obtain the best results of classification (Figure 7(a)), although these were slightly diminished in enhanced-resolution images (Figure 7(b)). This trend was observed in all wavelet analysis, excepting by applying the wavelet *Daubechies 20* (Figure 8). However, the performance of the method in this case did not overcome the results obtained with the wavelet *Biorthogonal 3.3* (Figure 7). Therefore, it is still difficult to suggest the resolution enhancement to improve the performance of the method. Probably, resolution enhancement via sparse representation should be more analyzed in future, increasing the resolution to larger

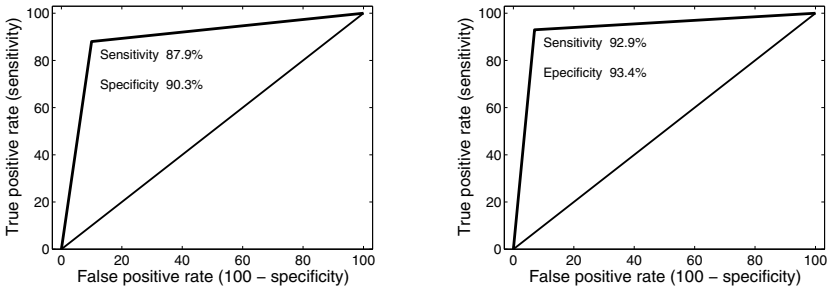


Fig. 8. ROC curves by applying the wavelet *Daubechies 20*, indicating the percentages of sensitivity and specificity in each case. a) non-enhanced resolution images. b) enhanced resolution images.

Table 1. Number of mammograms detected

	Biorthogonal 3.3		**Daubechies 20**	
	Non-enhanced resolution	Enhanced resolution	Non-enhanced resolution	Enhanced resolution
With microcalcifications	24/25	23/25	21/25	23/25
Without microcalcifications	287/297	284/297	269/297	278/297

rates in powerful computational media. Other algorithms for this purpose can be also tested [17], as well as the implementation of the method presented in this work in other mammograms databases. In any way, these results demonstrate the potential of this method to support the automated detection of microcalcifications in digital mammograms, including in those cases where the microcalcification spots are hard to detect by visual inspection (Figure 6).

References

[1] Nguyen, Fleiszer, D.: Interactive mammography analysis web tutorial. McGill Medicine. Molson Medical Informatics (2002), `http://sprojects.mmi.mcgill.ca/mammography/index.htm`

[2] Radiology - TIP, Xeromammography, Radiology - Technology Information Portal (2012), `http://www.radiology-tip.com/serv1.php?type=db1&dbs=Xeromammography`

[3] Cuttino, J.T., Yankaskas, B.C., Hoskins, E.O.L.: Screen film mammography versus xeromammography in the detection of breast cancer. British Journal of Radiology 59, 1159–1162 (2005), `http://bjr.birjournals.org/cgi/content/abstract/59/708/1159`

[4] AAPM, Equipment characteristics for xeromammography and screen-mammography. American Association of Physicists in Medicine (129) (1990)

[5] Nishikawa, R.M., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A.: Computer-aided detection of clustered microcalcifications on digital mammograms. Med. Biol. Eng. Comput. 33, 174–178 (1995)
[6] Gurcan, M.N., Yardimci, Y., Cetin, A.E., Ansari, R.: Detection of microcalcifications in mammograms using higher order statistics. IEEE Signal Process. Lett. 4, 213–216 (1997)
[7] Tang, J., Rangayyan, R.M., Xu, J., El Naqa, I., Yang, Y.: Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. IEEE Trans. on Inf. Tech. in Biom. 13, 236–251 (2009)
[8] Regentova, E., Zhang, L., Zheng, J., Veni, G.: Microcalcification detection based on wavelet domain hidden Markov tree model: Study for inclusion to computer aided diagnostic prompting system. Med. Phys. 34, 2206–2219 (2007)
[9] Jiang, J., Yao, B., Wason, A.M.: A genetic algorithm design for microcalcification detection and classification in digital mammograms. Comput. Med. Imag. Graph. 31, 49–61 (2007)
[10] Peng, Y., Yao, B., Jiang, J.: Knowledge-discovery incorporated evolutionary search for microcalcification detection in breast cancer diagnosis. Artif. Intell. Med. 37, 43–53 (2006)
[11] Suckling, J., Boggis, C.R.M., Hutt, I., Astley, S., Betal, D., Cerneaz, N., Dance, D.R., Kok, S.L., Parker, J., Ricketts, I., Savage, J., Stamatakis, E., Taylor, P.: The Mammographic Image Analysis Society Digital Mammogram Database. In: Excerpta Medica, MiniMammography Database. International Congress Series 1069, pp. 375–378 (1994),
`http://peipa.essex.ac.uk/ipa/pix/mias/miasdatabase.tar.gz`
[12] Yang, J., Wright, J., Huang, T., Ma, Y.: Image Super-Resolution via Sparse Representation. Trans. Img. Proc. 19, 2861–2873 (2010)
[13] Hou, H.S., Andrews, H.C.: Cubic spline for image interpolation and digital filtering. IEEE Transactions on Signal Processing 26, 508–517 (1978)
[14] Leiner, B.J., Lorena, V.Q., Cesar, T.M., Lorenzo, M.V.: Microcalcifications Detection System through Discrete Wavelet Analysis and Contrast Enhancement Techniques. In: Electronics, Robotics and Automotive Mechanics Conference, CERMA 2008, pp. 272–276 (2008)
[15] Pajares, J.G.: Visión por computador Imágenes digitales y aplicaciones, pp. 47–56. Alfaomega, Ra-Ma (2004)
[16] Laine, A.F., Schuler, S., Fan, J., Huda, W.: Mammographic feature enhancement by multiscale analysis. IEEE Trans. Med. Imag. 13, 725–740 (1994)
[17] Gutiérrez, R.M., Cerquera, E.A., Mañana, G.: MPGD for breast cancer prevention: a high resolution and low dose radiation medical imaging. Journal of Instrumentation 7, C07007 (2012), doi:10.1088/1748-0221/7/07/C07007

Influence of Interactions between Virtual Channels in Cochlear Implants

Ernesto A. Martínez-Rams[1] and Vicente Garcerán-Hernández[2]

[1] Universidad de Oriente, Avenida de la América s/n, Santiago de Cuba 90900, Cuba
eamr@fie.uo.edu.cu
[2] Universidad Politécnica de Cartagena, Antiguo Cuartel de Antiguones (Campus de la Muralla), Cartagena 30202, Murcia, España
vicente.garceran@upct.es

Abstract. Cochlear implants are devices which are surgically inserted into the internal ear, partially restoring hearing to people with profound deafness. Alternatives to improve levels of auditory perception are to increase the number of channels or frequency bands, that would need to increase the number of electrodes or electrode pairs per channel. Monopolar Virtual Channel (MPVC) stimulation mode increases the number of channels without increasing the number of electrodes using current steering. With this stimulation mode what still happens is interaction between the virtual channels, so some people who have these implants have not reported a striking improvement. In the present work, parameters of this electrical stimulation mode are modeled and evaluated. The results suggested that the maximal stimulation current can be determine so that no electrical interaction will occur between adjacent channels. If the stimulation current is higher than the calculated value, electrical interactions will occur between adjacent channels.

1 Introduction

Cochlear implants are electronic devices that which process sound to directly stimulate the auditory nerve fibers that innervate the cochlea which evokes the perception of hearing. These devices spectrally decompose each of the sonorous information in frequency bands. The energy information will be extracted to these bands. Each band has an electrical stimulation channel, which ends with an electrode or an electrode set.

At present, part of the new studies on these devices have the goal to improve neuronal electrical stimulation patterns and consequently increase the degree of auditory perception of people implanted. On the other hand, the study of various geometrical shapes of the electrodes leads to a better distribution of electric charge in the stimulated contact area. Similarly, the study of the electrodes array shape, has made it possible for the design of arrays capable of adapting to cochleae with congenital malformations. Furthermore, it is intended that these matrices become increasingly user-friendly for the surgical staff. Also, the study of the electrical stimulation mode seeks to improve the location and directionality

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 303–312, 2013.

of the current flow in the cochlea, this is the corresponding case outlined in this work. Both, monopolar and bipolar stimulation are electrical stimulation modes that have been used commonly in the cochlear implants. But new studies show the advantages of the tripolar and quadripolar stimulation.

In monopolar stimulation (MP), the current delivered by the active electrode, is received by the extra-cochlear electrode [1], as show in Figure 1. With an array of N electrodes is achieved stimulate N frequency bands. MP stimulation allows for greater selectivity and less interaction between channels with simultaneous stimulation, in the case of using the Simultaneous Analog Stimulation (SAS) strategy [2]. MP stimulation requires more high current than bipolar stimulation (BP) to achieve an increase of loudness. In the bipolar stimulation, the current delivered by the active electrode is received equally by the adjacent electrodes. With an array of N electrodes, it is possible to stimulate $N-1$ frequency bands and N bands when using medial-lateral electrode arranged in pairs. BP stimulation requires lower levels of stimulation current to achieve the same effects in loudness than MP stimulation. In both cases MP and BP, the stimulation of active electrode produces a broad current spreads and simultaneously an overlapping in the population of neurons, limiting the spectral information transmitted to the implanted people. In tripolar stimulation (TP), the current delivered by the active electrode is received equally by the adjacent electrodes, Figure 1. A variation of this stimulation is the Partial Tripolar (PTP), in wich the current supplied by the active electrode is received, a part and equally by the adjacent electrodes, and the other part by the extra-cochlear electrode Figure 1. With this stimulation it achieves some degree of current focusing which increases the loudness with increasing of the stimulation current level.

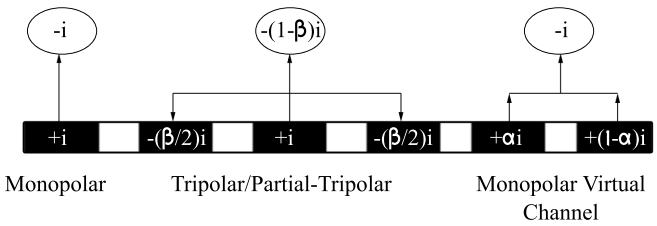


Fig. 1. Different stimulation modes. The oval above electrodes array represents the extra-cochlear electrode. Word "i"represents the current amplitude, α represents the fraction of current delivered to the basal electrode and β represents the fraction of current sent to the adjacent electrodes.

Monopolar Virtual Channel (MPVC) is another very interesting electric stimulator. Using current steering it achieves an increase to the number of virtual channels (VC) between two of active electrodes. This current is delivered by two adjacent active electrodes with the same or different current levels, and both are received by the extra-cochlear electrode. MPVC has been implemented by Advanced Bionics Harmony BTE processor which support the speech coding

strategy HiRes Fidelity 120 and Hires 90K implant family. In this, 120 frequency bands are divided into 120 virtual channels with only 16 electrodes. Some people implanted prefer the sound quality of Fidelity 120 [3,1], whereas others reported no significant advantage in the perception of sound in comparison with the HiRes strategy [3,4,1]. This is because MPVC strategy still has some interaction between channels due to fails in the current focusing. A study for a better use of this type of electrical stimulation is presented in our work.

2 Methods

2.1 Current Spread Model in MP Stimulation

For this work we use a simple model of current spread. With this model we can estimate the electrical excitation of the auditory nerves distributed throughout the cochlea when they are stimulated electrically with cochlear implants. The current density at a location of the cochlea is described as an exponential decay function [5], equation 1,

$$E\left(x(f)\right) = e^{\frac{-\left|X_{electrode} - x(f)\right|}{\lambda}} \tag{1}$$

where $x(f)$ corresponds to cochlea location, in mm from apex and frequency under analysis, whereas $X_{electrode}$ is the distance from apex of the active electrode, also in mm. Excitation spread is represented by λ, in mm. In the present work, the unit value is chosen in accordance with acoustic experiments performed with people implanted [5]. The equation 2 [6] specifies the cochlea location $x(f)$:

$$x(f) = log\left(k + \frac{f}{A}\right)\frac{L}{a} \tag{2}$$

where L is the human cochlea length (35 mm), and parameters A=165.4, a=2.1 and k=0.88 are constants.

2.2 Electrical Current Spread Model in MPVC Stimulation

The increase of the number of virtual channels is accomplished with current steering. To achieve this, an electrode is stimulates with a ratio α, the other is stimulated with a ratio $1 - \alpha$, over the total of stimulation current "i", Figure 1. α may have a value between 0 and 1. Some implanted patients show evidence of discrimination of $\Delta\alpha_{min}$=0.11 (equivalent to ten Virtual Channels).

When the value of α=0, all the current is delivered to apical electrode. Instead, when the value of α=1, all the current is delivered to the basal electrode. In both cases, the stimulation is equivalent to MP. In our work we use $\Delta\alpha$=0.25 (equivalent to five Virtual Channels per electrode pair).

For MPVC stimulation, the current density, in μA, at a section k of the cochlea can be represented as:

$$E_{MPVC}(\alpha, k) = I_{VC_{(\alpha)}} e^{\frac{-\left|X_{VC}(\alpha) - X_{section}(k)\right|}{\lambda}}; \alpha = 0...1; k = 1...K \tag{3}$$

where α is a coefficient used in the stimulation of this corresponding virtual channel. $X_{section}(k)$ is the location of the k section of the cochlea. $X_{VC}(\alpha)$ is the position between the two electrodes $n+1$ and n, of the corresponding virtual channel, and can be defined as follows:

$$X_{VC}(\alpha) = (1-\alpha)Xelectrode_{n-1} + \alpha Xelectrode_n \tag{4}$$

where $X_{electrode_n}$ is the location of the n electrode. Finally, $I_{VC_{(\alpha)}}$, is the corresponding VC current, and is obtained from equation:

$$I_{VC_{(\alpha)}} = I_{electrode_{n-1}} + I_{electrode_n} = I \cdot (1-\alpha) + I \cdot \alpha = I \tag{5}$$

2.3 Model for the Responses of Auditory-Nerve Fibers

The deterministic model for the responses of the auditory-nerve fiber (ANF) versus an electrical stimulation can be calculated by equation 6, [7]:

$$p(t) = \begin{cases} 0, \text{ for } V_{stim}(t) < V_{thr}(t) \\ 1, \text{ for } V_{stim}(t) \geq V_{thr}(t) \end{cases} \tag{6}$$

where $p(t)$ is discharge probability of the ANF at time t. $V_{stim}(t)$ is the stimulus level and $V_{thr}(t)$ is threshold level.

The stochastic model of the ANF discharge, when only a stimulus is present, is shown in equations 7 and 8 [8]. The level of noise is took account in this model.

$$p(t) = \begin{cases} 0, \text{ for } V_{stim}(t) < V_{thr}(t) + V_{noise}(t) \\ 1, \text{ for } V_{stim}(t) \geq V_{thr}(t) + V_{noise}(t) \end{cases} \tag{7}$$

where $V_{noise}(t)$ is the level noise of the Ranvier's node.

$$p(t) = \frac{1}{2}\left(1 + erf\left(\frac{V_{stim}(nt) - V_{thr}(t)}{\sqrt{2}\alpha}\right)\right) \tag{8}$$

The error function $erf(x)$ is twice the integral of the Gaussian distribution:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt \tag{9}$$

The stochastic model of the ANF discharge, when any stimulus is present, are shown in equations 10 and 11.

$$p(t) = \begin{cases} 0, \text{ for } V_{stim}(t) < V_{thr}(t) + V_{noise}(t) + V_{ref}(t) \\ 1, \text{ for } V_{stim}(t) \geq V_{thr}(t) + V_{noise}(t) + V_{ref}(t) \end{cases} \tag{10}$$

$$p(t) = \frac{1}{2}\left(1 + erf\left(\frac{V_{stim}(nt) - (V_{thr}(t) + V_{ref}(t))}{\sqrt{2}\alpha}\right)\right) \tag{11}$$

where $V_{ref}(t)$ is a reference level, which is function of the refractory period that represents the amount of time it takes for an neuron discharged to be ready for a second stimulus.

In this stochastic model, the discharge of the ANF in response to an electrical stimulus intensity is a Gaussian distribution [7], the probability density function is shown in equation 12, [9].

$$f(I,\mu,\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(I-\mu)^2}{2\sigma^2}} \tag{12}$$

where I is the electric intensity of stimulus, the parameter μ is the mean and σ^2 is the variance of the distribution.

The cumulative probability, is the total probability in the range -∞ to I, defined as:

$$P(I|\mu,\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{I} e^{\frac{-(I-\mu)^2}{2\sigma^2}} dI \tag{13}$$

In our simulations, we chose μ=50 $dB_{1\mu A}$ y σ=1 $dB_{1\mu A}$.

3 Results

In the MPVC method, electrical stimulation can increase N times the number of stimulation channels with respect to MP and BP stimulations. N depends on the value chosen for $\Delta\alpha$ and his minimum value differs from patient to patient.

A first test consisted in the simulation of stimulation current to two electrodes positioned at 14 mm and 15.2 mm from the apex, respectively. These electrodes have a separation equal to 1 bark that corresponds to frequencies 1000.0 Hz and 1192.4 Hz, respectively. Both electrodes are stimulated using a value α=0.5, achieving stimulate the position 14.57 mm, which corresponds to the virtual channel of 1092.5 Hz. Figure 2 a) shows the current density MPVC when both electrodes are stimulated and we concluded that using current steering, it can stimulate a specific area corresponding to a certain frequency of the cochlea and not just whole area between the two electrodes.

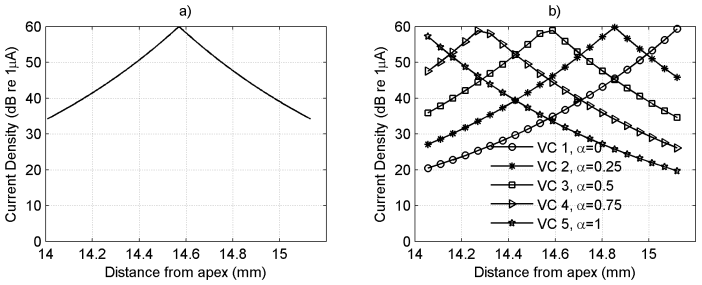


Fig. 2. Current density MPVC when two actives electrodes located at 14 and 15.2 mm from apex, are stimulated with I_{VC}=60 $dB_{1\mu A}$ and: a) α=0.5; b) α=0.00, 0.25, 0.50, 0.75, 1.00

In a second test, we simulated the stimulation of same electrodes but employing a value $\Delta\alpha$=0.25 now, to appreciate the possible electrical interaction between the five resulting virtual channels. Figure 2 b) shows that using five values of α (0.00, 0.25, 0.50, 0.75 and 1.00) the current steering of electrical stimulation is achieved, and resulting stimulation of neurons associated with 5 zones (Virtual Channels) of the cochlea between two electrodes located in central positions of 14.0080, 14.2890, 14.5700, 15.1320 and 14.8510 mm. According to E_{MPVC} curves of each virtual channel, there is an overlap in which the current density of a virtual channel may cause a neuronal firing to corresponding group of neurons adjacent to the virtual channel.

Knowing the probability of the neuron discharge as a function of applied electric stimulus intensity, Figure 3 and the results shown in Figure 2 b), we can investigate how to control the electrical interaction between virtual channels, which are shown in the following sections.

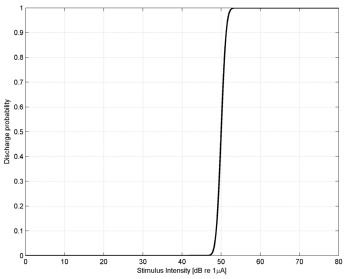


Fig. 3. Probability of the neuron discharge as a function of applied electric stimulus intensity

In all simulations, we chose: a threshold current I_{thr}=50 $dB_{re1\mu A}$, that corresponds to a probability of neuronal discharge equal to 50%. Standard deviation, σ=1 $dB_{re1\mu A}$. Mean value or threshold, μ=I_{trh}=50 $dB_{re1\mu A}$. α=0.00, 0.25, 0.50, 0.75, 1.00 and $\Delta\alpha$=0.25.

3.1 The Stimulation Current Density Is Greater Than the Threshold Current of Neuronal Discharging

Case 1: $E_{MPVC}(\alpha_i, k_n) = E_{MPVC}(\alpha_{i+1}, k_m) > I_{thr}; I_{VC} > I_{thr}$
The stimulation current I_{VC} of two adjacent virtual channels is above the threshold current I_{thr}. At the same time, the intersections of the current density curves of both adjacent virtual channels are greater than the threshold value, I_{thr}. In this case, there is a certain overlap of both curves so that the intensity of the electric stimulation of a virtual channel may produce the discharge of a cluster of neurons of adjacent virtual channel, which is interpreted as an electrical interaction between adjacent channels. The overlap area will be larger the higher

the stimulation currents of virtual channel. This at the same time produces a decrease in auditory perception of the implanted person.

Figure 4 a) shows the curves of five virtual channels that meet this condition. The black shaded area shows the area of overlap between the second and third virtual channels. Levels of less than I_{thr} curves $= 50\,\mathrm{dB}_{1\mu A}$ not taken into account by considered zero or near zero their discharge probability, i.e. neuronal discharges hardly occur.

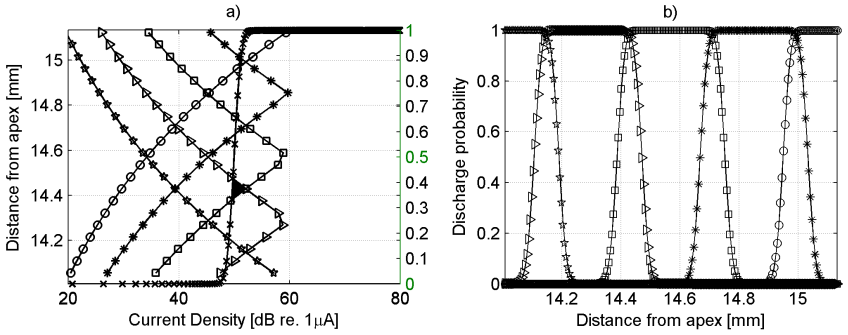


Fig. 4. In the case 1, I_{VC}=60 $dB_{re1\mu A}$. a) Cochlear zone vs MPVC and probability of the neuron discharge vs electric stimulus intensity; b) Probability of the neuron discharge vs cochlear zone.

Case 2: $E_{MPVC}(\alpha_i, k_n) = E_{MPVC}(\alpha_{i+1}, k_m) = I_{thr}; I_{VC} > I_{thr}$
The stimulation current I_{VC} of two adjacent virtual channels is above the threshold current I_{thr}. At the same time, the intersections of the current density curves of both adjacent virtual channels are equal to the threshold value, I_{thr}. In this case there will be no overlap of both curves, so that the intensity of the electrical stimulus of a virtual channel may not produce the discharge of a group of neurons of adjacent virtual channel. This is interpreted as no electrical interaction will occur between adjacent channels. From the psychophysical point of view it can be interpreted as an optimal auditory perception of the implanted person, Figure 5.

The stimulation current I_{VC} value, in which occurs that both current density curves are equal to the current threshold value, I_{thr}, should be equal to the maximum current of stimulation to use between the pair of electrodes n and $n+1$ to not produce electrical interaction between adjacent channels. That is, higher levels of current will occur interactions between channels, which is undesirable in cochlear implants.

Equation 14 is used to calculate the maximum stimulation current per virtual channel to prevent electrical interaction to occur between virtual channels. This assumes that all channels are stimulated with maximum current I_{VC}. X_o represents the arithmetic mean between adjacent channels, and X_{VC} the virtual channel position.

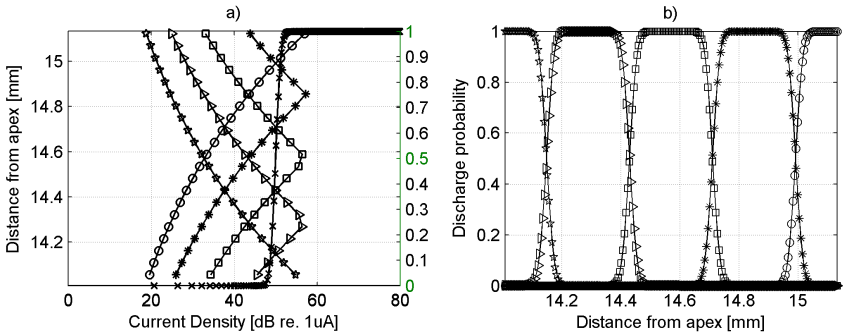


Fig. 5. In the case 2, I_{VC}=57.5426 $dB_{re1\mu A}$. a) Cochlear zone vs MPVC and probability of the neuron discharge vs electric stimulus intensity; b) Probability of the neuron discharge vs cochlear zone.

$$I_{vc_{max}} = I_{thr} e^{\frac{+|x_{VC} - x_o|}{\lambda}} \tag{14}$$

Case 3: $E_{MPVC}(\alpha_i, k_n) = E_{MPVC}(\alpha_{i+1}, k_m) \leq I_{thr}; I_{VC} \geq I_{thr}$
The stimulation current I_{VC} of two adjacent virtual channels is still higher to the threshold current I_{thr}. At the same time, the intersections of the current density curves of both adjacent virtual channels are lower than the threshold value, I_{thr}. As in the case 2, there is no overlapping of both curves, so that the intensity of the electrical stimulus of a virtual channel may not produce the discharge of a group of virtual channel adjacent neurons. This is interpreted as non-electrical interaction between the adjacent channels, Figure 6.

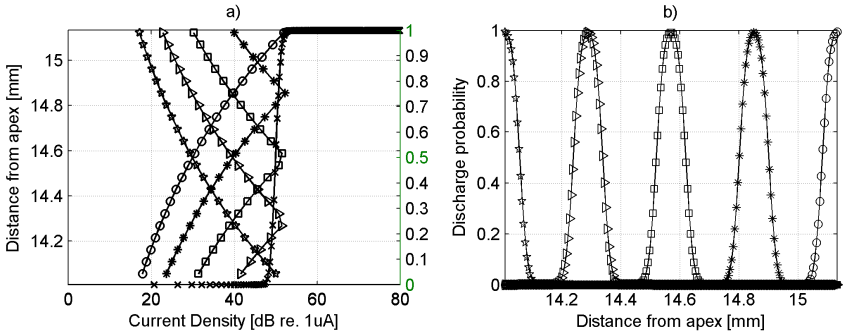


Fig. 6. In the case 3, I_{VC}=55.5426 $dB_{re1\mu A}$. a) Cochlear zone vs MPVC and probability of the neuron discharge vs electric stimulus intensity; b) Probability of the neuron discharge vs cochlear zone.

3.2 The Stimulation Current Density Is Lower Than the Threshold Current of Neuronal Discharging

Case 4: $E_{MPVC}(\alpha_i, k_n) = E_{MPVC}(\alpha_{i+1}, k_m) < I_{thr}; I_{VC} < I_{thr}$
The stimulation current I_{VC} of two adjacent virtual channels is lower to the threshold current I_{thr}, the intersections of the current density curves of both adjacent virtual channels are lower than the threshold value, I_{thr}. In this case neuronal discharge will not occur in any of the virtual channels. From the psychophysical point of view, it can be interpreted as absence/very low sound levels that cannot evoke perception, Figure 7.

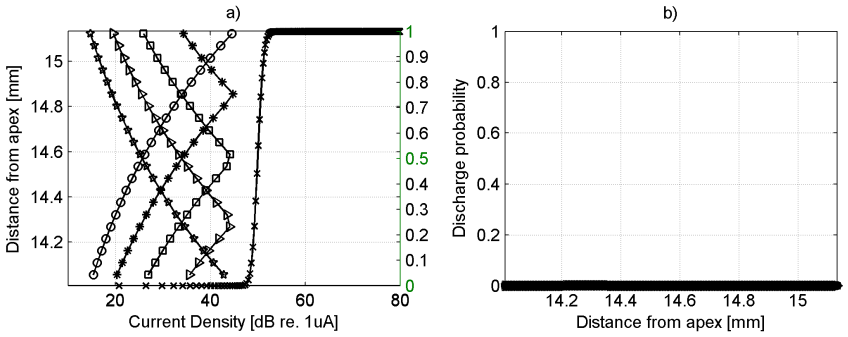


Fig. 7. In the case 4, I_{VC}=45 $dB_{re1\mu A}$. a) Cochlear zone vs MPVC and probability of the neuron discharge vs electric stimulus intensity; b) Probability of the neuron discharge vs cochlear zone.

4 Conclusions

Using current steering permits improvement of the number of stimulation channels in cochlear implants. One clear example of this is the stimulation MPVC. However, its performance can be increased by computing the value of maximum stimulation current per virtual channel to prevent electrical interaction to occur between virtual channels. This value is a function of the neural discharge threshold. The higher the level of electrical stimulation with respect to the maximum value calculated per channel, the greater the area where electrical interactions occur between channels. If the stimulation current is lower to the threshold current, the neuronal discharge will not occur in any of the virtual channels. Therefore, we can conclude that dynamic range value of electrical stimulation must be between the threshold value and the maximum previously calculated. In other words, levels near the threshold will evoke neural discharge of a very narrow area, which from the psychophysical point of view -although not shown in this work-, would be perceived as a decrease in the intensity of the sound (Loudness). Instead, electrical stimulation levels near to the maximum value will evoke the neural discharge of an entire area of the virtual channel,

which is perceived as increase of loudness. Electrical stimulation levels greater than maximum value, will produce electrical interaction between channels which would not be beneficial for people with cochlear implants.

Acknowledgements. This research has been partially funded by the Thematic Network "Computacin Natural y Artificial para la Salud (CANS) of CYTED, Spain. The authors also acknowledge to Biomedical Processing Signal Project of "Universidad de Oriente, Cuba.

References

1. Srinivasan, A.G., Landsberger, D.M., Shannon, R.V.: Current focusing sharpens local peaks of excitation in cochlear implant stimulation. Hearing Research 270, 89–100 (2010)
2. A. B. Corporation, CLARION TM. Multi Strategy Cochlear Implant. V2.0., 12740 San Fernando Road Sylmar, California 91342-3728 (1996)
3. Brendel, M., Buechnear, A., Krueger, B., Frohne Buechner, C., Lenarz, T.: Evaluation of the Harmony soundprocessor in combination with the speech coding strategy HiRes 120. Otol Neurotol. 29, 199–202 (2008)
4. Berenstein, C.K., Mens, L.H.M., Mulder, J.J.S., Vanpouke, F.J.: Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations. Ear Hear 29, 250–260 (2008)
5. Laneau, J., Moonen, M., Wouters, J.: Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. Journal Acoustical Society of America 119(1), 491–506 (2006)
6. Greenwood, D.D.: A cochlear frequency-position function for several species-9 years later. Journal Acoustical Society of America 87, 2592–2605 (1990)
7. Bruce, I.C., White, M.W., Irlicht, L.S., O'Leary, S.J., Dynes, S., Javel, E., Clark, G.M.: A Stochastic Model of the Electrically Stimulated Auditory Nerve: Single-Pulse Response. IEEE Transaction on Biomedical Engineering 46(6), 617–629 (1999)
8. Bruce, I.C., Irlicht, L.S., White, M.W., O'Leary, S.J., Dynes, S., Javel, E., Clark, G.M.: A Stochastic Model of the Electrically Stimulated Auditory Nerve: Pulse-Train Response. IEEE Transaction on Biomedical Engineering 46(6), 630–637 (1999)
9. Bruce, I.C., White, M.W., Irlicht, L.S., O'Leary, S.J., Clark, G.M.: The Effects of Stochastic Neural Activity in a Model Predicting Intensity Perception with Cochlear Implants: Low-Rate Stimulation. IEEE Transaction on Biomedical Engineering 46(12), 1393–1404 (1999)

Dealing with Bullying through Genetic Algorithms

M. Angélica Pinninghoff[1], Pedro L. Salcedo[2], Ricardo Contreras[1], Andrea Yáñez[1], and Eduardo Oportus[1]

[1] Department of Computer Science
University of Concepción, Chile
[2] Research and Educational Informatics Department
University of Concepción, Chile
{mpinning,psalcedo,rcontrer}@udec.cl

Abstract. Aggression in schools is a problem for which there is no a simple solution. On the other side, it is known that a specific configuration in the distribution of students can affect the behavior among them. Based on a previous experience, we propose to apply genetic algorithms in order to deal with the large number of configurations that can arise on these types of problems. Introducing the concept of penalization has shown to be an interesting concept that allows to reach feasible solutions in reduced computing times. Real environments were considered to conduct the experiments. The set of solutions has been analyzed an accepted as a helpful tool to minimize negative interactions in a classroom.

Keywords: Genetic algorithms, Bullying, Sociogram.

1 Introduction

It is not unusual to read reports about aggressive behavior of children in a classroom. This kind of violence, known as bullying [5,16], is today a common practice in many schools. It is clear that abuse (physical or psychological), aggression and violence among students, can arise in most of schools. Teachers, educational managers, and school psychologists tell us that, very frequently, this kind of behavior reach the public knowledge status when aggression and victimization is a reality and its effects are dramatic.

Bullying is a product of bad interpersonal relationships among students. This issue shows the influence that the relationships network generated in a specific group has, a situation that can be strengthen by the social environment in which the group is located.

Additionally, it is well known that in every group, a leader arises. And the leader has a key role in aggression and victimization processes, and in the environment that acts as a social network, because he can have a direct influence in teaching-learning processes. In words of Brown [2], *what characterize a leader is that they can influence in others more than they can be influenced by others.* A previous definition remarks the key role of a leader: the leader in a group is the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 313–322, 2013.

individual with higher authority, understanding it as the capability of a person to influence a great number of members in a regular way.

It is independent if the leader is chosen by a group or imposed by an external authority; his particular behavior produces different effects on the group members. A positive leader will improve the group productivity, motivation and group cohesion, while a negative leader will use charismatic tendencies to manipulate, isolate, and ostracize his or her followers.

Cerezo et al. [4] have developed a set of techniques for allowing early bullying detection. Among these techniques, the Bull-S Test is designed to evaluate aggression in students [3], which is introduced as a valid tool for detecting, measuring and quantifying aggressiveness. This test has three main objectives: i) to allow the analysis of socio-affective characteristics in a group of classmates, ii) to help teachers in the process of detecting abuse behavior and iii) to generate intervention proposals.

On the other side, the Ohio group has developed a series of questionnaires for measuring leadership. The first instrument the group designed to measure leader behavior is the Leader Behavior Description Questionnaire (LBDQ) [9,6,15], in which a group of persons is asked to describe their leader, typically a supervisor, in terms of the frequency the leader shows behaviors described in the questionnaire.

The way in which people interact has been a research area for many years, as shown in [1], where concepts as connectivity are used to describe properties such as distance among individuals or to describe, alternatively, the number of different relationships that it is possible to establish between two individuals. Models about the way in which ideas spread through a social network, are considered in [11], where authors show efficient algorithms to identify, in a social network, those individuals who have a higher degree of influence. This work is focused on medical and technological innovations.

The work of Jackson [10] offers a comprehensive insight for different vision under which a network can be observed, and presents different ways for handling classic concepts as friendship, proposing interpretations on how a specific structure (of a network) can affect a particular behavior, introducing measures for social interactions.

A simple and direct mechanism to represent social networks is to use graphs. There is an important number of experiences, as in the work of [13], that studies popular Internet sites as YouTube and Flickr, that identify users as individuals that belong to social networks and uses graphs as a tool that improves curent systems by adding new applications.

The work we present here is based on a recent experience that identifies specific students distribution in a classroom, to diminish the negative leadership effects; in other words, to reduce bullying [14]. This experience offers a starting point for applying artificial life - based algorithms, working on a population of classroom configuration, for finding optimal configurations in terms of influences. Our proposal is an improved version that considers different possibilities for distributing students, different configurations for genetic algorithms, and a

set of tests carried out on real school environments. Inclusion of penalization, as a factor which take into account undesirable configurations, is a relevant element in this work.

This article is structured as follows. The first section is made up of the present introduction; the second section describes the problem we are dealing with; the third section presents a resume on genetic algorithms, while the fourth section shows our proposal. Results are shown in fifth section, and the final section is devoted to conclusions of the work.

2 The Problem

In simple words, the problem we are dealing with, is to minimize the negative influence some students have on their classmates, in a classroom. The hypothesis is that it is possible to find a particular distribution of students, in such a way that a potential victim is located in a place not easily reachable by a potential aggressive student.

The key is to build a sociogram, which gives us information about aggressive relationships, obtained by taking sets of tests carried out by the teachers [14]. A sociogram is a set of indexes and graphics obtained from procedures developed for analyzing intra group relationships; aimed to describe and measure the structures of relationships underlying in small groups. The sociogram is the input data for a genetic algorithm that search for optimal distributions in a classroom. An optimal distribution is one in which no potential victim is in the neighborhood of a potential aggressive classmate.

The sociogram is represented by using a graph, in which nodes are connected through lines if there is a relationship between these points. In other words, the graph is a useful mechanism to represent binary relationships, i.e., relationships that allow to describe some type of connection between two individuals.

In formal terms, a graph is a pair of sets, V and E, where V represents the sets of points and E represents the existing relationships between two particular points. If G = (V,E) is the graph, then E = V × V [8]. It is said that an element v in V is a node and an element e in E is an edge. Some interesting properties can emerge in such a representation. For example, it is accepted that the relationship between two individuals is symmetric if we can say that there is a specific relationship between an individual A and an individual B, and that it is equivalent to say that there is a specific relationship between an individual B and an individual A. A typical example of such a relationship is friendship. Not every relationship present this property; in aggression, we can say that A is victim of B, but it is not equivalent to say that B is victim of A.

A typical graph looks like the one shown in figure 1 (taken from [14]) in which nodes represent students, and edges represent relationships among them. As it can be seen, there are some nodes with a higher degree of influence on their neighbor nodes. This is the first step to identify interesting individuals.

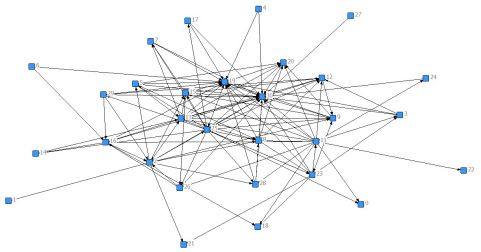


Fig. 1. A typical graph showing relationships among students

3 Genetic Algorithms

Genetic algorithms (GA) are a particular class of evolutionary algorithms, used for finding optimal or good solutions by examining only a small fraction of the possible space of solutions. GAs are inspired by Darwin's theory about evolution. The basic concept of GAs is designed to simulate processes in natural system necessary for evolution, specifically those that follow the principles of survival of the fittest. As such they represent an intelligent exploitation of a random search within a defined search space to solve a problem.

The structure of a genetic algorithm consists of a simple iterative procedure on a population of genetically different individuals. The phenotypes are evaluated according to a predefined fitness function, the genotypes of the best individuals are copied several times and modified by genetic operators, and the newly obtained genotypes are inserted in the population in place of the old ones. This procedure is continued until a *good enough* solution is found [7].

In this work, a chromosome represents a specific distribution of student in a classroom. In genetic algorithm terminology, each location is a gen and the identification of a student represents the value (allele) that the gene has. A good fitness means that a particular distribution contains a minimum number of potential victims (zero, ideally) in the neighborhood of an aggressive student.

Genetic Operators: Different genetic operators were considered for this work. These genetic operators are briefly described bellow:

- Selection. Selection is accomplished by using the roulette wheel mechanism [7]. It means that individuals with a best fitness value will have a higher probability to be chosen as parents. In other words, those configurations that are not a good distribution of students in a classroom are less likely selected.
- Cross-over. Cross-over is used to exchange genetic material, allowing part of the genetic information that one individual has, to be combined with part of the genetic information of a different individual. It allows us to increase genetic variety, in order to search for better solutions. Due to the nature of the problem we are dealing with, specific cross-over operators are considered: OX and PMX [12].

– Mutation. By using this genetic operator, a slight variation is introduced into the population so that a new genetic material is created.

4 Our Proposal

To illustrate our proposal, let us consider two groups, that represent different features. Let us call them *class1* and *class2*. In *class1* aggressive relationships are distributed on the classroom; in *class2*, aggressive relationships are concentrated in a small group. Students belonging to both groups are similar in ages (from 14 to 16 years old). *Class1* consists of 30 students, and each one of them considers that is a victim or aggressor of, at least, one classmate. *Class2* consists of 36 students; only 3 of them are considered aggressors, and there are some students that do not see themselves as a victim of a classmate.

This proposal considers three classroom configurations. These configurations represent physical distributions as shown in Figure 2. The distribution in the first group shows a classroom with no aisles (type 1); students are distributed homogeneously, in individual chairs. The distribution in the second group corresponds to a classroom that has an aisle for each two students (type 2). Finally, the distribution in the third group shows a schema of a classroom in which every student has an aisle on each side (type 3).

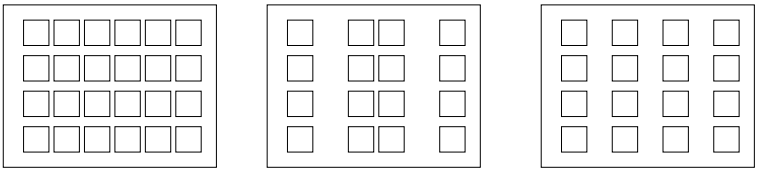


Fig. 2. Types of classroom

A classroom with their students is represented as a matrix of integers. Every entry in the matrix contains an integer that identifies a particular student, and the location in the matrix corresponds to the location in the classroom. Aisles are denoted by a zero value. A chromosome is a one dimension matrix, which size depends on the number of students, and contains a concatenation of rows in the original matrix that represents the classroom. In genetic terms, a chromosome is a permutation of integers, taking into account that are some locations that are not to be changed. These fix values correspond to aisles.

Initial population is randomly generated. A distribution of students is generated and then, in a random process, it is generated an exchange of students. This process is repeated for generating each chromosome belonging to the initial population.

The evaluating function (fitness), allows to evaluate the quality of each distribution (chromosome). A lower value for the fitness means a higher chromosome quality. Evaluation function is shown in equation (1); where N is the number of students, $Influence(x_i)$ is the number of potential victims pointing to student x in location i, $Neighborhood(i)$ is a value assigned to neighbors of i, that is considered in equations (2, 3 and 4) depending on the type of classroom (see figure 3); Loc is a factor that represents the distance between a particular student and the teacher. The higher the value, the grater the distance.

Figure 3 illustrates the effect that different types of neighborhood have. The black square denotes the location under consideration. The variable α denotes a direct neighborhood (horizontal or vertical neighborhood). The variable β represents a close diagonal neighborhood. The variable γ represents a direct neighborhood, but in which students are separated by an aisle, and the variable δ represents a diagonal neighborhood through an aisle. Finally, $Penalty(x_i)$ is a factor to penalize the fact that a potential victim and a potential aggressor are neighbors.

$$Fitness = \sum_{i=l}^{N} (Influence(x_i) * Neighborhood(i) + Penalty(x_i)) \qquad (1)$$

Equation (2) computes the value associated to the first configuration shown in Figure 3, which corresponds to the neighborhood of classroom type 1.

$$Neighborhood(i) = Loc * (\alpha_1 + \ldots + \alpha_4 + \beta_1 + \ldots + \beta_4) \qquad (2)$$

Equation (3) computes the value associated to the second configuration shown in Figure 3, which corresponds to the neighborhood of classroom type 2.

$$Neighborhood(i) = Loc * (\alpha_1 + \ldots + \alpha_3 + \beta_1 + \beta_2 + \gamma + \delta_1 + \delta_2) \qquad (3)$$

Equation (4) computes the value associated to the third configuration shown in Figure 3, which corresponds to the neighborhood of classroom type 3.

$$Neighborhood(i) = Loc * (\alpha_1 + \alpha_2 + \gamma_1 + \gamma_2 + \delta_1 + \ldots + \delta_4) \qquad (4)$$

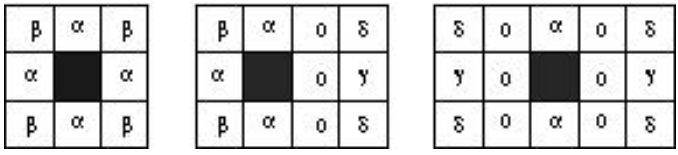


Fig. 3. Influence of neighbors on a specific location

Penalty is computed as follows: if a potential aggressor is located in a neighborhood that corresponds to an α label, then fitness is increased in a value a; if the location corresponds to a β label, then fitness is increased in a value b. A

similar procedure associates γ to c and δ to d. If there is no potential aggressor in the neighborhood, penalty is zero. Equations (5, 6 and 7) express the penalty which corresponds to neighborhoods shown in Figure 3. Table 1 shows the values considered for a, b, c and d.

$$Penalty(x_i) = a_1 + \ldots + a_4 + b_1 + \ldots + b_4 \tag{5}$$

$$Penalty(x_i) = a_1 + \ldots + a_3 + b_1 + b_2 + c + d_1 + d_2 \tag{6}$$

$$Penalty(x_i) = a_1 + a_2 + c_1 + c_2 + d_1 + \ldots + d_4 \tag{7}$$

Table 1. Values used for neighborhood penalization

Parameters	Values
a	30, 40, 100, 200, 250, 400, 600
b	20, 75, 100, 150, 200, 300
c	10, 50, 75, 100, 150
d	5, 25, 50, 100

5 Results

Experiments were carried out by considering three classroom configurations (type 1, type 2 and type 3) and two classes (*class1* and *class2*). In *class1*, the sociogram shows an important number of aggression relationships. Every student plays a role as a victim or as an aggressor. In *class2*, it is possible to identify only 3 aggressors; in this class, most of students do not behave as victims or aggressor.

The goal of experiments is to show that it is possible to obtain a distribution of students in a classroom, in such a way that it is minimized the number of relationships involving a potential victim or a potential aggressor in a neighborhood.

Table 2 shows parameters used in the genetic algorithm. Due to the fact that the genetic algorithm was implemented by using a toolbox provided by Matlab, the mutation percentage is always a value that complements the crossover percentage. Two crossover operators (OX and PMX) were implemented, because classic operators take a first part from one of the parents and the second part from a different parent, to give birth to a new chromosome. The problem with this approach is that for two parents, a specific student can appear on the first part of parent one, and in the second part of the second one, generating that the same student will appear in a offspring twice. For mutation, two alleles randomly selected are interchanged, and the selection of individuals was accomplished by using roulette wheel.

Table 2. Values used for genetic parameters

Parameters	Values
Population	200, 250, 300
Number of generations	200
Crossover (percentage)	50, 60, 70, 80, 90, 100
Mutation (percentage)	50, 40, 30, 20, 10, 0

Table 3. Values used for neighborhood quantification

Parameters	Values
α	1, 2
β	1, 2
γ	1
δ	0.5

Table 3 shows values considered for testing, that take into account the effect of neighbors.

Results show that it is possible to obtain, in all cases, perfect solutions, i.e., configurations in which the number of aggression relationships is zero. Table 4 summarizes the results of the conducted experiments; it shows the percentage of perfect solutions obtained.

It can be seen that for classroom type 1 (the more complex distribution of chairs, with no aisles), and taking into account *class1*, we obtained a 13% of perfect solutions by considering a population size of 200 individuals; 47% of perfect solutions by considering a population size of 250 individuals, and 60% by considering a population size of 300 individuals. If we take into account *class2*, we obtained a 94%, for every considered population size. For classroom type 2 and classroom type 3, interpretation of results is analog.

Table 4. Summary showing percentage of perfect solutions

		Population Size		
		200	250	300
Classroom type 1	class 1	13	47	60
	class 2	94	94	94
Classroom type 2	class 1	20	20	47
	class 2	100	89	89
Classroom type 3	class 1	30	40	50
	class 2	100	90	80

6 Conclusions

Artificial life offers an interesting alternative in order to deal with problems in which the number of situations to consider is large enough to make it difficult the traditional approach.

In this experience, the inclusion of penalization allows to obtain solutions in no more than 200 generations. Without penalization, a stable solution is reached after 1000 generations, and a larger population size. This element, not considered in previous works, has shown a positive impact on solutions.

Additionally, experiments show that population size is not a relevant element. On the other side, best solutions arise when values used for penalization are larger.

Finally, a reduced population and a small number of genetic generations allow to rapidly obtain satisfactory results.

References

1. Barnes, J.A.: Graph Theory and Social Networks: A Technical Comment on Connectedness and Connectivity. Sociology 3(2) (1969)
2. Brown, R.J.: Group proccess: Dynamics within and between groups. Blackwell, Oxford (1988)
3. Cerezo, F.: El Test Bull-S Instrumento para la evaluación de la agresividad entre escolares. Albor-Cohs, Madrid (2000)
4. Cerezo, F., Calvo, A., y Sánchez, C.: Bullying y estatus social en el grupo aula en una muestra de escolares. In: IV Congreso Internacional de Psicología y Educaci; Almería (2004)
5. Cerezo, F.: Violencia y Victimización entre escolares. El Bullying: estrategias de identificación y elementos para la intervención a través del Test Bull-S. Revista Electrónica de Investigación Psicoeducativa. 4(9), 333–352 (2006)
6. Fleishman, E.A.: The measurement of leadership attitudes in industry. Journal of Applied Psychology 37, 153–158 (1953)
7. Floreano, D., Mattiussi, C.: Bio-Inspired Artificial Intelligence. Theories, Methods, and Technologies. The MIT Press (2008)
8. Gross, J.L., Yellen, J.: Handbook of Graph Theory Discrete Mathematics and Its Applications. CRC Press (2003)
9. Hemphill, J.K.: Leader behavior description. Ohio State Univerity, Personel Research Board, Columbus (1950)
10. Jackson, M.: Social and Economic Networks. Draft. Princeton University Press (2008)
11. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C (2003)
12. Michalewicz, Z., Fogel, D.: How to Solve it: Modern Heuristics. Springer (2000)
13. Mislove, A., et al.: Measurement and analysis of online social networks. In: Internet Measurement Conference. Proceedings of the 7th ACM SIGCOM Conference on Internet Measurement, San Diego, California (2007)

14. Pedro Salcedo, L., Angélica Pinninghoff, M., Ricardo Contreras, A.: Group formation for minimizing bullying probability. A proposal based on genetic algorithms. In: Ferrández, J.M., Álvarez Sánchez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2011, Part II. LNCS, vol. 6687, pp. 148–156. Springer, Heidelberg (2011)
15. Stogdill, R.M., Coons, A.E.: Leader behavior: Its description and measurement. Ohio State University, Bureau of Business Research, Columbus (1957)
16. Sutton, J., Smith, P.K.: Bullying as a group process: An adaptation of the participant role approach. Aggressive Behavior 25, 97–111 (1999)
17. Willging, P.: Técnicas para el análisis y visualización de interacciones en ambientes virtuales. Redes - Revista Hispana Para el Análisis de Redes Sociales 14(6) (Junio 2008)

Using Ant Colony Optimization for Edge Detection in Gray Scale Images

Ricardo Contreras, M. Angélica Pinninghoff, and Jaime Ortega

Department of Computer Science
University of Concepción, Chile
{mpinning,rcontrer,jaimefortega}@udec.cl

Abstract. Digital image processing is a research topic that has been studied from different perspectives. In this paper we propose an approach based on a paradigm that arises from artificial life; more specifically ant colonies foraging behavior. Ant based algorithms have shown interesting results in a wide variety of situations, including optimization and clustering. In this work we compare different ant colony algorithms on a set of images, for the detection of edges. Results are presented as images, in which ants have built specific solutions, and discussed within an image-processing framework.

Keywords: Edge detection, ACO Algorithms.

1 Introduction

Digital images are two-dimensional representations of images, a set of pixels arranged in a matrix, in order to allow an electronic manipulation. In a gray-scale, each pixel is represented by using a numeric value belonging to the interval [0,255], that illustrates different shades of gray, varying from black at the weakest intensity to white at the strongest. On the other way, segmentation [9], [13] is an operation which divides an image into parts or regions having a particular feature. Segmentation algorithms are based on two intensity-based properties; the first one is the similarity among pixels, while the second one focuses on continuity related to gray levels, the last approach involves a family of algorithms for detecting edges by considering strong intensity changes among neighbors pixels. In recent years, Ant Colony Optimization (ACO) [6] algorithms have been developed to detect images' edges, by taking inspiration from the ants behavior [1], [7], [11], [16], [17].

Some ant families have the capability of finding the shortest path between their nest and the source of food. Ants use the environment as a medium for communication. They exchange information indirectly by depositing pheromone, while they pass through a particular trial (or path). The information exchanged has a local scope, only an ant located where the pheromone were left has a notion of them. This system property is called *stigmergy* and occurs in many social animal societies (it has been studied in the case of the construction of pillars in the nests of termites). The mechanism to solve a problem too complex

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 323–331, 2013.

to be addressed by only one ant is a good example of a self-organized system. This system is based on positive feedback (the deposit of pheromone attracts other ants that will strengthen it themselves) and negative feedback (dissipation of the route by evaporation prevents the system from thrashing). Theoretically, if the quantity of pheromone remained the same over time on all edges, no route would be chosen. However, because of feedback, a slight variation on an edge will be amplified allowing thus the choice of an edge. The algorithm will move from an unstable state, in which no edge is stronger than another, to a stable state where the route is composed of the strongest edges. The basic philosophy of the algorithm involves the movement of a colony of ants through the different states of the problem influenced by two local decision policies, viz., trails and attractiveness. Thereby, each such ant incrementally constructs a solution to the problem. When an ant completes a solution, or during the construction phase, the ant evaluates the solution and modifies the trail value on the components used in its solution.

The above described behavior is the inspiration source for using artificial ants [6], aimed to solve optimization problems.

The idea of using artificial ants to solve hard problems has been developed by different authors. In [11] is proposed an ACO algorithm hybridized with 2-OPT for fractal image compression. In [10] ACO algorithms are used in image segmentation, improving thresholding algorithms. Thresholding algorithms are the focus in [10], authors obtain experimental results to demonstrate that the proposed ant-based method performs better than other two established thresholding algorithms.

Related to previous work, closer to our proposal, in [1] authors use an ACO algorithm for image edge detection. Edge detection is accomplished by seeking pixels that show important differences with respect to their neighbors, in terms of intensity level (in gray-scale).

The work described in [17] presents an approach that obtained interesting results. They utilize a number of ants moving on a 2-D image for constructing a pheromone matrix, each entry of which represents the edge information at each pixel location of the image. The movements of the ants are steered by the local variation of the image's intensity values.

In [16] authors propose an ant colony optimization based algorithm for continuous optimization problems on images like image edge detection, image compression, image segmentation and structural damage monitoring in image processing. They show the feasibility of the algorithm in terms of accuracy and continuous optimization. This work emphasizes an important feature: a good solution, like the shortest path, has more pheromone than the longest paths.

The objective of this work is to evaluate the effectiveness of different ACO algorithms, in edge detection. In particular, this article focuses in gray-scale images. Our approach differs from the works in [7], [1] and [17], because they consider only one Ant Colony algorithm, and we take into account a set of two algorithms: Ant System (AS) and Elitist Ant System (EAS). Additionally, we compare the obtained results with the deterministic *Canny* procedure.

This article is structured as follows; the first section is made up of the present introduction; the second section describes the images (RAPD). The third section describes the ACO algorithms we used. The fourth section shows the results we obtained, and the final section shows the conclusions of the work.

2 RAPD Images

Randomly Amplified Polymorphism DNA (RAPD) images are grey scale images composed by lanes and bands, which has been used in verifying genetic identity. The lanes are the vertical columns shown in Figure 1 and each one of them represents a DNA sample, except the reference lanes which are the leftmost and the rightmost lanes.

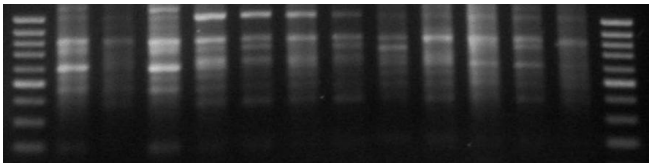


Fig. 1. A sample RAPD image with two reference lanes

The reference lanes are used to indicate the molecular weight, measured in base pairs (bp), of the DNA. The bands are the horizontal lines in each lane that represent the segments agglomeration of a DNA sample with the same bp value (see Figure 2).

Fig. 2. A sample of a lane in a RAPD image

The process of producing RAPD images is affected by many physical-chemical factors [14], and generates different kind of noise, deformations and diffusion, among others, in the lanes and bands of the RAPD images. To correct these problems is important, because their effects can lead to erroneous biological conclusions.

The RAPD image correction is a research field not completely explored. There are software tools like ImageJ [3], Gel Compar II[1], GelQuant [5] and Quantity One [2] used for RAPD image analysis. However, the correction is done manually by an expert so the abnormalities remain undetected to the human eye. This fact is pointed out in [14] where the authors use the Radon transform to correct the whole image, and mathematical morphology with cubic spline smoothing for band detection. On the other hand, in [15] is proposed a polynomial algorithm to support Multilevel Thresholding. [8] adopts that polynomial multilevel thresholding algorithm to identify the bands in a lane, however it was not possible to get a successful detection in band location. This problem acts as a reason for finding a mechanism to detect the bands, and in doing so we propose the use of different ACO algorithms.

3 ACO Algorithms

In this work, two algorithms which consider ACO were implemented. They are based on [7] and the first one employs an Ant System (AS) algorithm and the second one corresponds to an Elitist Ant System (EAS). They are based on the AS algorithm [6], and they are adapted for the specific problem.

In the AS algorithm, artificial ants walk through a space of solutions represented as a graph $G = (V, E)$, where V is the set of nodes and E is the set of edges or connections between nodes. Edges of the graph are the places where ants deposit pheromone. For implementing this model, the space of solutions is represented by a matrix of pixels of an image. Each pixel is a matrix entry, an edge in the graph represents a neighborhood between two pixels.

In ACO algorithms, an arbitrary number of ants are randomly distributed on the matrix pixels. When an ant walk through the image, it deposits pheromone. The amount of pheromone depends on the contrast among neighbor pixels. High contrasts implies more pheromone. This is reflected in the matrix, by increasing the amount of pheromone in those pixels that present some degree of intensity contrast with respect to the previous pixel in an ant movement. A higher contrast implies a higher amount of pheromone on the destiny pixel.

In this scheme, a solution is a configuration that results from the fact that every pixel in the original image has been traversed, and different regions in the image present different gray-scale intensities, due to different pheromone deposits. When a solution is obtained, it is followed (in practice, after a time interval) by a process in which the pheromone is diminished, representing the effect of the time on the pheromone that was deposited in the different paths. This process, besides to model closely a real phenomena occurring with ants, allows to avoid the effect of local minimum.

In this approach, the amount of pheromone T deposited by an ant k on each edge is given by equation (1), where η and p are constants and Δ_{gl} is the difference computed as the median intensity among the previous node and its neighbors, and the median intensity among the current node and its neighbors. Finally

[1] Details are available at `http://www.appliedmaths.com/gelcompar/gelcompar.htm`

q is a constant value, fixed in 255, that represents the maximum brightness for a pixel. It allows to deposit more pheromone in higher contrast regions in the image.

$$T_k = \eta + \frac{p\Delta_{gl}}{q} \tag{1}$$

The neighborhood of a pixel i comprises the 8 pixels (nodes) around pixel i in an image. When the pheromone is deposited, the ant moves to a new position, in a probabilistic way, taking into account the path traversed by the ant and the amount of pheromone (the intensity) in neighbors cells. The probability that an ant located at node (cell) i moves to node j is shown in equation (2) where two weights are considered: $W(\tau_{i,j})$ described in equation (3), that represents the impact of pheromone, where $(\tau_{i,j})$ is the amount of pheromone associated to the path (i, j), and $w(dir)$ that determines the weight associated to the direction that an ant can take, depending on the previous path that this ant has traversed.

Figure 3 illustrates a set of values associated to the probability, of choosing a particular path.

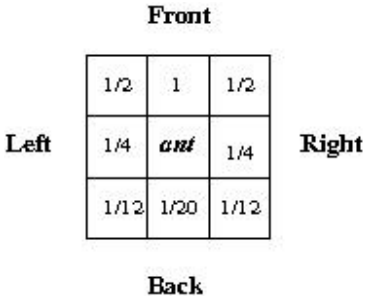


Fig. 3. Values associated to different possible paths

For $w(dir)$ we used the values suggested in [7]: $w(Front) = 1$; $w(Front - Diagonal) = 1/2$; $w(Right, Left) = 1/4$; $w(Back - Diagonal) = 1/12$; and $w(Back) = 1/20$.

Where *Front* denotes that an ant walks straightforward, keeping the previous direction in a path, *Front-Diagonal* denotes a diagonal path that combines a right (or left) movement with a front movement and *Back-Diagonal* denotes a diagonal path that combines a right (or left) movement with a back movement; *RightLeft* denotes a movement that changes the current direction to the left or to the right, and *Back* indicates the ant chooses go back, to the previous step in its traversal. It reflects the fact that the highest value is assigned to the ant walking straightforward, and the lowest value is assigned when the ant chooses to move going back through the path that has been just traversed.

$$P_{ij} = \frac{W(\tau_{i,j})w(dir)}{\sum_{l/j\epsilon N_i} W(\tau_{i,l})w(dir)} \tag{2}$$

N_i denotes the set of neighbors nodes for node i.

$$W(\tau_{i,j}) = (1 + \frac{\tau_{i,j}}{1 + \delta \cdot \tau_{i,j}})^{\beta} \quad (3)$$

In equation 3, β is a constant used to control the attraction effect that the pheromone has on the ants; δ is another constant that indicates the capability of ants to detect the pheromone.

EAS [6] differs from AS in the fact that there is a reinforcement of pheromone in edges corresponding to the best solution found in one iteration. In this case, the pheromone reinforcement is accomplished by taking into consideration the path traversed by the the ant that detects the highest differences of intensity. The amount of deposited pheromone is given by equation 1.

Canny algorithm is a well known method used for edge detection. Even though is quite old, it has become one of the standards edge detection methods [4]. The steps in the Canny edge detector are as follows: smooth the image with a two dimensional Gaussian; take the gradient of the image; non maximal suppression and edge thresholding [12].

Given the characteristics of the Canny algorithm, it is interesting to compare the results using this algorithm with the results obtained using ants. Specifically, we compare results using Canny with results using AS and EAS.

4 Results

For testing we used a set of jpg images, in gray scale, having different sizes, measured in pixels.

In AS and EAS algorithms, we considered different values for three parameters: initial population, pheromone evaporation rate, and the value that considers the variable pheromone deposit (q in equation 1). Besides that, some tests were carried out to determine the number of iterations required to obtain the maximum contrast in an image.

The other parameters were selected based on the work of [7], and these are the following: β=3.5, δ=0.2, η=0.07, p= 1.5.

In figures 4, 5 and 6 we show the images that presented the best results.

For all images: a) original image; b) image processed with AS algorithm, 500 iterations, population size equal to 30% of size image (in pixels), pheromone evaporation rate equal to 0.015; c) image processed with EAS algorithm, 1100 iterations, population size equal to 30% of size image (in pixels), d) Canny.

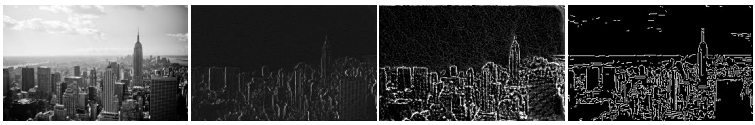
According to results, the best tested parameter combination corresponds to the image processed with the EAS algorithm, after 600 iterations, population size of 30% with respect to the original image and pheromone evaporation rate equal to 0.015. The difference can be produced by the fact that in this approach the pheromone is reinforced in the edges corresponding to the best solution for each iteration, in other words, in those places in which there is a high intensity change between pixels. It means that the pheromone concentrates on edges from the beginning of the process. On the other side, the other considered algorithms encourage, in general, the search of alternative paths.

(a) Lena (b) AS (c) EAS (d) Canny

Fig. 4. Tests with Lena

(a) Cat (b) AS (c) EAS (d) Canny

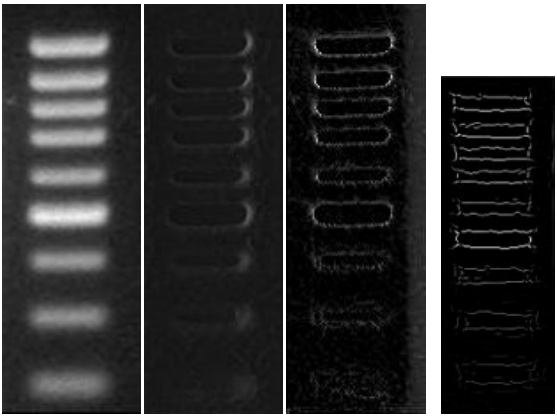
Fig. 5. Tests with cat



(a) New York (b) AS (c) EAS (d) Canny

Fig. 6. Tests with New York images

Results obtained with RAPD images were worse than expected, due to the low quality of some images. This issue made necessary an image pre-processing. During this pre-processing the contrast was increased and the image was binarized. To binarize an image allows us to obtain less fuzzy regions for helping in edge detection. In an ordinary gray-scale image, the process of detecting edges can led to non precise results. Some of obtained images are shown in Figure 7, where a) original image; b) image processed with AS algorithm, 500 iterations, population size equal to 30% of size image (in pixels) pheromone evaporation rate equal to 0.015; c) image processed with EAS algorithm, 600 iterations, population size equal to 30% of size image (in pixels), pheromone evaporation rate equal to 0.015.



(a) lane (b) AS (c) EAS (d) Canny

Fig. 7. Tests with RAPD images

In some cases, when dealing with low contrast images, some details are not detected by ants (see the image for New York city in figure 6). Ant-based algorithms cannot detect clouds in the picture, but Canny can do it. In the case of *Lena* (figure 4), where we can observe a well defined gray-scale shade, ant-based algorithms get lines with a high degree of continuity, and detect edges that can't be detected using Canny; in particular in the hair of Lena. In the case of the image for *Cat* (figure 5), we find some difficulties, because of the white color that makes edges difficult to detect.

With RAPD images, in spite of obtaining a poor result, it was possible for ant-based algorithms to detect all bands present in the image. Canny allows to detect all bands too, but edges are not clear, as it can be seen in figure 7.

For this kind of images it is necessary to continue with the experimental process, searching for new parameter combinations.

In general, EAS algorithms exhibited acceptable results, due to the fact that there is a pheromone reinforcement in candidate solutions generated by the ants. Anyway, it is important to find an equilibrium point to reflect the trade-off between the pheromone evaporation rate and the amount of pheromone an ant deposits. As with many parameters it must be done experimentally.

5 Conclusions

In this paper we have carried out a series of experiments involving ants as a mechanism for edge detection. It is clear that results obtained are improved when there is a reinforcement of pheromone for promising solutions. Shade differences are also important when detecting edges, with a uniform light distribution, as in the case of the image with the cat, the process of detecting edges is not particularly successful.

RAPD images are complex, end even the use of a standard method cannot guarantee good results, particularly due to the lack of contrast in the available images. To take into account, as a future work, we suggest to test new pre-processing methods, before applying ant-based algorithms.

References

1. Baterina, A.V., Oppus, C.: Image Edge Detection Using Ant Colony Optimization. International Journal of Circuits, Systems and Signal Processing 10(2) (2010)
2. Bio Rad Laboratories. Quantity One User's Guide. Bio-Rad Laboratories (2000)
3. Burger, W.: Digital Image Processing with Java and ImageJ. Springer (2006)
4. Canny, J.: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8(6) (November 1986)
5. Das, R., Laederach, A.: GelQuant User's Manual. Stanford University (2004)
6. Dorigo, M., Stuetzle, T.: Ant Colony Optimization. The MIT Press (2004)
7. Fernandes, C.M.: A Camera Obscura for Ants. In: Sigevolution, Newsletter of the ACM Especial Interest Group on Genetic and Evolutionary Computation, vol. 3, pp. 9–16 (2008)
8. Carolina Gárate, O., Angélica Pinninghoff, J.M., Ricardo Contreras, A.: Band Correction in Random Amplified Polymorphism DNA Images Using Hybrid Genetic Algorithms with Multilevel Thresholding. In: Ferrández, J.M., Álvarez Sánchez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2011, Part II. LNCS, vol. 6687, pp. 426–435. Springer, Heidelberg (2011)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Pearson/Prentice Hall, Upper Saddle River (2008)
10. Malisia, A., Tizhoosh, H.: Image Thresholding Using Ant Colony Optimization. In: The 3rd Canadian Conference on Computer and Robot Vision, CRV 2006 (2006)
11. Martinez, C.: An ACO algorithm for image compression. CLEI Electronical Journal 9(2), paper 1 (2006)
12. Nabernehad, E., Sharifzadeh, S., Hassanpour, H.: Edge Detection Techniques: Evaluation and Comparisons. Applied Mathematical Sciences 2(31), 1507–1520 (2008)
13. Phillips, D.: Image Processing in C, 2nd edn. R & D Publications (1994)
14. Rueda, L., Uyarte, O., Valenzuela, S., Rodriguez, J.: Processing Random Amplified Polymorphism DNA Images Using the Radon Transform and Mathematical Morphology. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 1071–1081. Springer, Heidelberg (2007)
15. Rueda, L.: An Efficient Algorithm for Optimal Multilevel Thresholding of Irregularly Sampled Histograms. In: da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C., Loog, M. (eds.) S+SSPR 2008. LNCS, vol. 5342, pp. 602–611. Springer, Heidelberg (2008)
16. Kaur, S., Agarwal, P., Rana, R.S.: Ant Colony Optimization: A Technique used for Image Processing. International Journal of Computer Science and Technology 2(2) (June 2011)
17. Tian, J., Yu, W., Xie, S.: An ant colony optimization algorithm for image edge detection. IEEE Congress on Evolutionary Computation (2008)

Modeling the Effect of Fixational Eye Movements in Natural Scenes

Andrés Olmedo-Payá[1], Antonio Martínez-Álvarez[2], Sergio Cuenca-Asensi[2], José Manual Ferrández-Vicente[3], and Eduardo Fernández[1]

[1] Institute of Bioengineering and CIBER BBN,
University Miguel Hernandez, Alicante, Spain
e.fernandez@umh.es
[2] Computer Technology Department, University of Alicante,
Carretera San Vicente del Raspeig s/n, 03690 Alicante, Spain
amartinez@dtic.ua.es
[3] Department of Electronics and Computer Technology,
Universidad Politécnica de Cartagena, Spain
jm.ferrandez@upct.es

Abstract. Our eyes never remain still. Even when we stare at a point, small involuntary movements move our eyes in an imperceptible manner. Researchers agree on the presence of three main contributions to eye movements when we fix the gaze: *microsaccades*, *drifts* and *tremor*. These small movements carry the image across the retina stimulating the photoreceptors and thus avoiding fading. Nowadays it is commonly accepted that these movements can improve the discrimination performance of the retina. In this paper, several retina models with or without fixational eye movements were implemented by mean of *RetinaStudio* tool to test the feasability of these models to be incorporated in future neuroprosthesis. For this purpose each retina model have been stimulated with the same natural scene sequence of images. Results are discussed from the point of view of a neuroprosthesis development.

Keywords: Fixational eye movements, microsaccades, retina model.

1 Introduction

Our eyes are always in constant motion. Some of these movements are involuntary and appear even when we fix our gaze. Although some of these movements are relatively large and displace the image across the retinal photoreceptors, they are not perceptible to us [1].

Since the late 1800s, several research groups have been investigating in understanding the role of these fixational eye movements in the vision [2] using a variety of techniques for recording [3]. Nowadays, the most accepted the idea about the role of these movements is that they can improve discrimination performance in ways not explicable just by prevention of visual fading [4] [5] [6] [7]. In particular, *microsaccades* are probably the eye movements with the greatest potential to perform this task [8] [9] [10].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 332–341, 2013.

Various approaches have been proposed to study fixational movements. On one hand, Ditchburn et al. [11], Nachmias et al. [12] and others used recording techniques to proof the role of these eye movements in visual perception. On the other hand, simulation of *microsaccades* were used to asses in a realistic manner the role of eye movements in the primary retinal responses. Greschner et al. [13] in their work, performed an experiment which demonstrated that the simulated small movements, like *tremor*, in a turtle retina, activated sets of retinal ganglion cells in a synchronized manner. Finally, the approach based on retina models has several advantages from the point of view of experimentation. Comparing to an *in vivo* experimental setup, retina models can be easily modified and used as often as desired without needed of laboratory animals or annoying human testing. Moreover retina models increases the range of possibilities allowing, for example, describe disease retinas, describe different properties in isolation and more options that could not be done with real retinas.

In this way Donner and Hemilä [14] attempted to clarify the effects of these movements on the messages that retinal cells send to the brain by mean of mathematical models or Martinez et al. [15] and Wohrer et al. [16], whose work was focused on the development of new retina models.

Our group is working on the development of a cortical visual neuroprosthesis aimed to restore some functional vision to profoundly visual-impaired people. For this purpose, the objective of the current study is to test the feasibility of fixational eye movements to be implemented in a visual neuroprosthesis. In this way, different retina models with and without fixational eye movements have been described and tested to check the vision improvement. Results show that retina models including eye movements have a better behavior than models without this feature.

2 Retina Model

The retina plays an important role in visual perception of humans. It is responsible for converting the outside world images into electrical signals understandable by the visual cortex of the brain. In fact, it is considered as a part of the brain. This process must be unequivocal and fast enough to ensure recognition of objects within a few milliseconds [17]. Therefore a good retina model, as well as its physical implementation, should take into account this time constraint to be able to respond to stimuli in real time.

In this paper several retina models which are sensitive to variations in luminance are described by mean of *RetinaStudio* [18], a framework to encode visual information that allows the specification, testing, simulation, validation and implementation of bioinspired retina models. Following the work of Ferrandez et al. [19] which uses a *Utah Electrode Array* [20] to perform extracellular recordings, each retina model is defined as a 10×10 matrix of different kind of ganglion cells. Thereby, models of *On* ganglion cells and *Off* ganglion cells, with and without fixational eye movements are described making a total of four different retina models:

- ***On* model**, consisting in only 10×10 *On* ganglion cells.
- ***On + eye movements* model**, consisting in 10×10 *On* ganglion cells with eye movements.
- ***Off* model**, consisting in only 10×10 *Off* ganglion cells.
- ***Off + eye movements* model**, consisting in 10×10 *Off* ganglion cells with eye movements.

RetinaStudio allows us to describe the different stages that comprises a biological retina. The first stage, inspired in the Outer Plexiform Layer of the retina, was modeled spliting the images into three color channels (R,G and B).

The second stage, inspired in the Inner Plexiform Layer, was modeled by means of spatial filters, more specifically with a network of the well known *Difference of Gaussian* (*DoG*) filter (see equation 1).

$$DoG(x, y, \sigma_1, \sigma_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \quad (1)$$

where σ_1 and σ_2 take the values 0.9 and 1.2 respectively in the *On* retina models and take 1.2 and 0.9 values in *Off* retina models. The magnitude value of these parameters have already been contrasted in the work of Morillas et al. [21]. The *Difference of Gaussian* filter receives contributions of the three types of photoreceptors R, G and B (for red, green and blue), and thus generates a mexican-hat contribution for every color channel. This contribution simulates the circular shape and the antagonist center-periphery behavior of the receptive fields of ganglion cells in the retina, see Fig. 1.

Finally, inspired in the Ganglion Cell Layer, the leakage-Integrate&Fire spiking neuron model proposed by Gerstner and Kistler [22] is used to model the ganglion cell behavior.

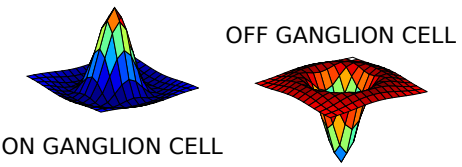


Fig. 1. Center-periphery behavior of *On* and *Off* receptive fields

All retina models represent a piece of fovea having a size of 1.8×1.8 *mm*, where each receptive field is about 180μm of diameter [23].

2.1 Eye Movements

The three main contributions of involuntary eye movements were integrated within our retina models: *tremor*, *drifts* and *microsaccades*.

Tremor is the amplitude-smallest of all eye movements. It consists in an aperiodic oscillation of the eyes with small amplitudes and frequencies within the range of the recording system's noise [24]. Following the work of Ratliff and Riggs [3], these movements are described in our model as oscillatory waves with an small amplitude, in a range between 2,83μm and 4,13μm and with a frequency of about 50Hz. The oscillatory waves generated are superimposed on *drifts*, see Fig. 2.

Drifts are slow curvy motion that occurs between *microsaccades* and appear simultaneously with *tremor*. This movement can displace the image across a dozen photoreceptors with a mean speed of 6 $min\ s^{-1}$ [2]. *Drifts* are described using the *gamma* distribution of equation 2 as proposed by [25]. In our case, $\lambda = 1$ and different k values are randomly chosen taking integer values between 1 and 9. The angle of these *drifts* is randomly modified using equations 3, taken into account that all *drifts* must be directed outwards [26].

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{(k-1)!} \tag{2}$$

$$\begin{aligned} x' &= xcos(\theta) - ysin(\theta) \\ y' &= xsin(\theta) + ycos(\theta) \end{aligned} \tag{3}$$

In our model each drift movement have a duration between 0.24s and 0.48s, with a mean of 0.36s, and an amplitude between 68μm and 100μm, with a mean of 0.84μm. The values of duration and amplitude were randomly chosen within these ranges.

Microsaccades are fast eye movements of short duration, about 25ms [4], displacing the image across a range of several dozen to several hundred photoreceptors widths. The role of *microsaccades* in visual perception have been debated for years. The most accepted idea is that its main role is to prevent fading and thus keep the vision [11] [12].

All *microsaccades* are modeled in our retina model as rectilinear movements directed to the center of the visual scene and appearing just after *drifts* with a random angle and amplitude. The amplitude of each *microsaccade* have a mean of 100μm and a duration of 20ms. Fig. 2 shows an example of fixational eye movements paths. As can be seen, high-frequency oscillatory paths (*tremor*) are superimposed on *drifts* (curved lines) forming only one path moving slowly. *Microsaccades* (rectilinear lines) appear just after *drifts* and move the image across the retina quickly.

3 Experiments and Results

To assess the role and behavior of *microsaccades* and the other small eye movements in the perception of natural scenes, each retina model (*On*, *On + eye*

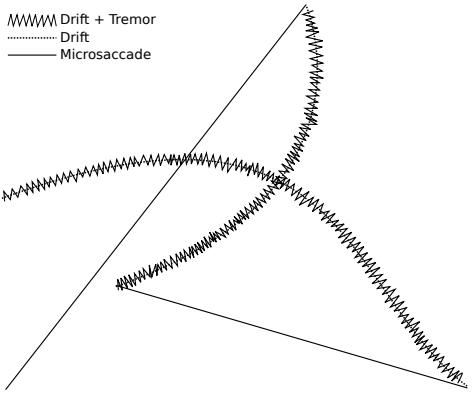


Fig. 2. *Microsaccades*, *drifts* and *tremor* movements. Continuous line shows the fast *microsaccade* movement. Dashed line shows the slow curve of *drifts* movements. Oscillatory line shows the combination of *drifts* and *tremor*.

movements, *Off* and *Off + eye movements*) was stimulated using a stimulus of 8,5 seconds of duration from a video file. The stimulus consists in a 25 fps video with a resolution of 622×622 pixels and 8-bits of colour depth (gray-scale). The video was created using *Visionegg* [27]. The records were generated in *Neural Events File* format [28] and processed with *NeurALC* [29]. This test was repeated 10 times.

Top plot in the Fig. 3 shows the whole natural scene, where a superimposed black rectangle indicates the image area stimulating our retina model at a given time. This rectangle travels across the image from left to right during the stimulation. As a measure of retinal activity it was considered the *gray mean value* [30] (equation 4), represented in the plot as color-pale graphic and calculated from the natural scene as:

$$M(x) = \frac{1}{m} \sum_{y=1}^{m} (x, y) \tag{4}$$

where $n \times m$ is the image resolution and (x, y) represent a pixel.

Middle and bottom plots show, with black points, the spike trains produced by 100 electrodes in a raster plot. The *population activity* graphic is represented by a dark line and the *gray value mean* is superimposed in both plots for comparison purposes.

Inspecting the figure, it can be seen that the activity is greater in the lighter areas and lower in the dark areas of the image as expected. The electrodes located on the lighter areas have greater activity than those located in other areas (see raster plot). Moreover, the *population activity* graphic is more similar to *gray value mean* graphic in the retina model with eye movements implemented than those without them. It should be noted that the *population activity* have been normalized to values between 0 and 255 for comparison.

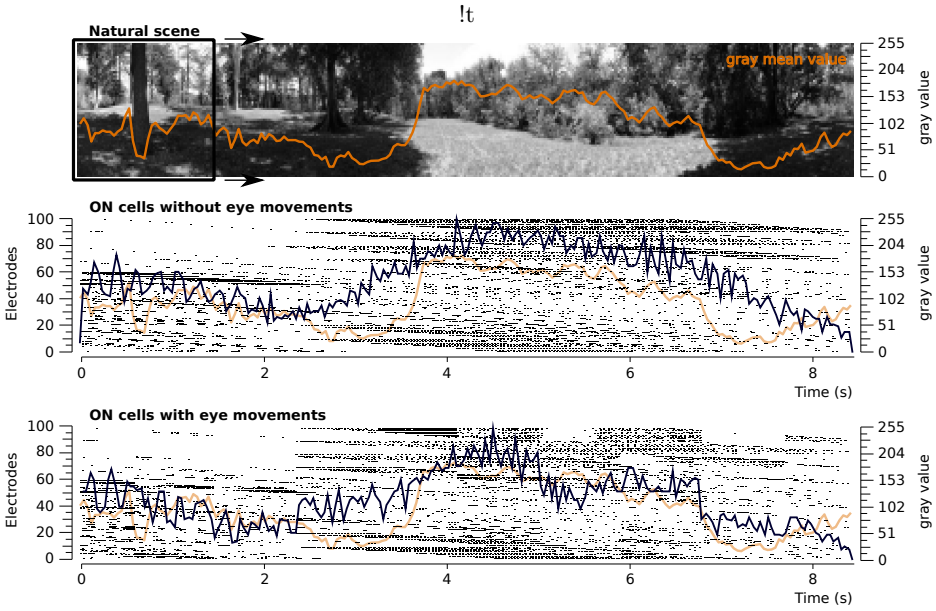


Fig. 3. Raster plot showing the spike trains produced for 100 electrodes and *population activity* with a bin of $40ms$ for *On* cells

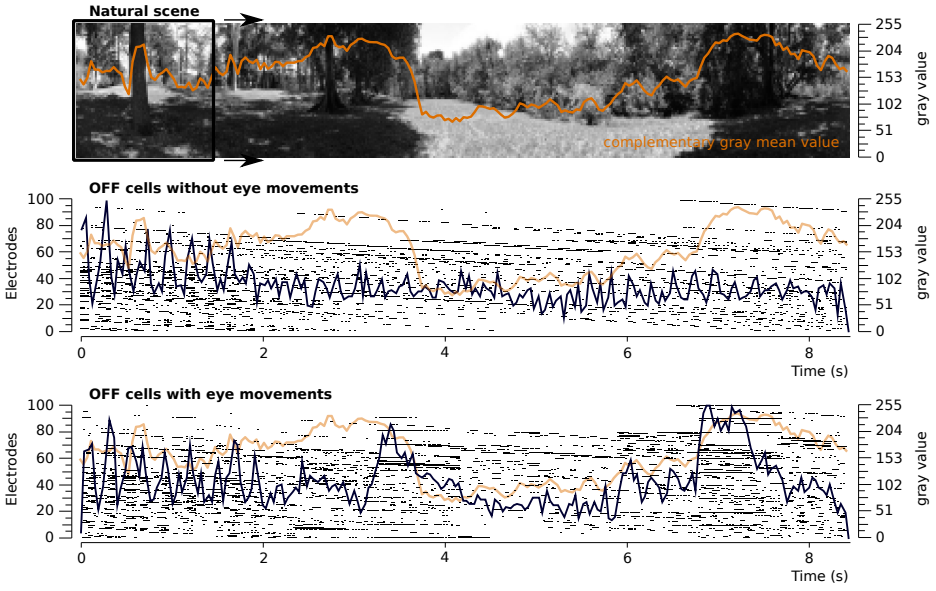


Fig. 4. Raster plot showing the spike trains produced for 100 electrodes and *population activity* with a bin of $40ms$ for *Off* cells

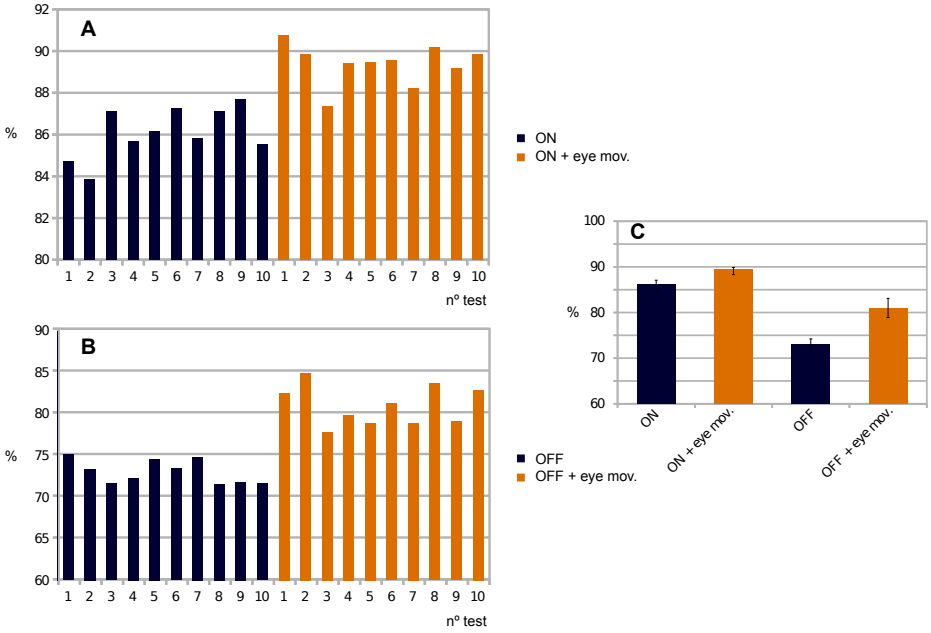


Fig. 5. Percent similarity between *population activity* and *gray value mean*. **A** Similarity in ten simulations for *On* model with and without eye movements. **B** Similarity in ten simulations for *Off* model with and without eye movements.**C** Average of ten stimulations in *On* model and *Off* model with and without eye movements.

Fig. 4 is similar to Fig. 3 but shows *population activity* of *Off* retina model compared with the complementary *gray value mean*, see equation 5.

$$M'(x) = 255 - \frac{1}{m}\sum_{y=1}^{m}(x, y) \tag{5}$$

In this case, unlike Fig. 3, the activity is greater in the dark areas and lower in the lighter areas of the image. When comparing with the *population activity*, it is clear that both signals are closer when eye movements are applied.

To measure how different is the *population activity* of our models when applying or not eye movements, a similarity measure based in a subtraction was performed between *gray value mean* graphic and *population activity* graphic. It is showed in equation 6,

$$S_{12} = \frac{1}{N}\sum_{n=1}^{N}(x_1[n] - x_2[n])^2 \tag{6}$$

where $x_1[n]$ and $x_2[n]$ corresponds to *gray value mean* and *population activity* respectively.

The similarity measure provided by this subtraction is depicted in Fig. 5, where the 100% corresponds to identical response in the *gray value mean* and the retina model ($S_{12} = 0$ in equation 6). Plots **A** and **B** show the similarity of ten stimulations of the four retina models, plot **C** shows the average of ten stimulations. Pale bars corresponds to retina models without eye movements and dark bars corresponds to retina models with eye movements.

As can be seen, there is a remarkable difference between dark bars and pale bars, being closest to *gray value mean* the pale bars. In the average graphic, plot **C**, the difference is about 4% in the *On* models and 8% in the *Off* models.

In general, the similarity is greater when the retina model includes in its implementation fixational eye movements.

4 Conclusions

The role of fixational eye movements in natural scenes using retina models has been assessed in this paper. For this purpose the main three classes of involuntary eye movements (*tremor*, *drifts*, and *microsaccades*) have been implemented in two ganglionar retina models: *On* model and *Off* model. Results have been compared using the same models without eye movements implemented.

The results of this comparison indicate a significant improvement in the visual perception in natural scenes when retina models include *microsaccades*, *tremor* and *drift* movements. Furthermore, neuronal activity records show greater sensitive to light changes, improving the edge recognition. These results indicate that it is possible to improve environment perception if these small movements are incorporated in the retina models and encourage us to implement these eye movements in a future visual neuroprosthesis. The comparison of these models with biological records will be performed as a future work.

References

1. Martinez-Conde, S., Macknik, S.L., Hubel, D.H.: The role of fixational eye movements in visual perception. Nature Reviews Neuroscience 5, 229–240 (2004)
2. Yarbus, A.L.: Eye Movements and Vision. Plenum, New York (1967)
3. Ratliff, F., Riggs, L.A.: A involuntary motions of the eye during monocular fixation. J. Exp. Psychol. 40, 687–701 (1950)
4. Ditchburn, R.W.: The function of small saccades. Vision Res. 20, 271–272 (1980)
5. Gerrits, H.J., Vendrik, A.J.: Artificial movements of a stabilized image. Vision Research 10, 1443–1456 (1970)
6. Krauskopf, J.: Effect of retinal image motion on contrast thresholds for maintained vision. Journal of the Optical Society of America 47, 740–744 (1957)
7. Sharpe, C.R.: The visibility and fading of thin lines visualized by their controlled movement across the retina. Journal of Physiology 222, 113–134 (1972)

8. Steinman, R.M., Cunitz, R.J., Timberlake, G.T., Herman, M.: Voluntary control of microsaccades during maintained monocular fixation. Science 155, 1577–1579 (1967)
9. Zuber, B.L., Stark, L.: Saccadic suppression: Elevation of visual threshold associated with saccadic eye movements. Experimental Neurology 16, 65–79 (1966)
10. Riggs, L.A., Ratliff, F.: The effects of counteracting the normal movements of the eye. Journal of the Optical Society of America 42, 872–873 (1952)
11. Ditchburn, R.W., Fender, D.H., Mayne, S.: Vision with controlled movements of the retinal image. J. Physiol. 145(1), 98–107 (1959)
12. Nachmias, J.: Determiners of the Drift of the Eye during Monocular Fixation. J. Opt. Soc. Am. 51, 761–766 (1961)
13. Greschner, M., Bongard, M., Rujan, P., Ammermüeller, J.: Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. Nature Neuroscience 5, 341–347 (2002)
14. Donner, K., Hemila, S.: Modelling the effect of microsaccades on retinal responses to stationary contrast patterns. Vision Research 47(9), 1166–1177 (2007) ISSN 0042-6989, 10.1016/j.visres.2006.11.024
15. Martínez, A., Reyneri, L.M., Pelayo, F.J., Romero, S.F., Morillas, C.A., Pino, B.: Automatic Generation of Bio-inspired Retina-Like Processing Hardware. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 527–533. Springer, Heidelberg (2005)
16. Wohrer, A., Kornprobst, P.: Virtual Retina: A biological retina model and simulator, with contrast gain control. Journal of Computational Neuroscience 26, 219–249 (2009)
17. Bongard, M., Ferrandez, J.M., Fernandez, E.: The neural concert of vision. Neurocomputing 72, 814–819 (2009)
18. Martìnez-Álvarez, A., Olmedo-Payá, A., Cuenca-Asensi, S., Ferrández, J.M., Fernández, E.: RetinaStudio: A bioinspired framework to encode visual information. Neurocomputing (2012) ISSN 0925–2312, 10.1016/j.neucom.2012.07.035
19. Ferrández, J.M., Bongard, M., García de Quirós, F., Bolea, J.A., Ammermüller, J., Normann, R.A., Fernández, E.: Decoding the Population Responses of Retinal Ganglions Cells Using Information Theory. In: Mira, J., Prieto, A.G. (eds.) IWANN 2001. LNCS, vol. 2084, pp. 55–62. Springer, Heidelberg (2001)
20. Normann, R., Maynard, E.M., Rousche, P., Warren, D.: A neural interface for a cortical vision prosthesis. Vision Res. 39, 2577–2587 (1999)
21. Morillas, C.A., Romero, S.F., Martínez, A., Pelayo, F.J., Fernández, E.: A Computational Tool to Test Neuromorphic Encoding Schemes for Visual Neuroprostheses. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 510–517. Springer, Heidelberg (2005)
22. Gerstner, W., Kistler, W.: Spiking Neuron Modelssingle Neurons, Populations, Plasticity, pp. 93–97. Cambridge University Press (2002)
23. Gauthier, J.L., Field, G.D., Sher, A., Greschner, M., Shlens, J., Litke, A.M., Chichilnisky, E.J.: Receptive Fields in Primate Retina are Coordinated to Sample Visual Space More Uniformly. PLoS Biol. 7(4), 1000063 (2009), doi:10.1371/journal.pbio.1000063
24. Carpenter, R.H.S.: Movements of the Eyes. Journal of Modern Optics 36(9), 1273–1276 (1989)
25. Stacy, E.W.: A Generalization of the Gamma Distribution. Ann. Math. Statist. 33(3), 1187–1192 (1962)

26. Pritchard, R.M.: Stabilized images on the retina. Sci. Am. 204, 72–78 (1961)
27. Straw, A.D.: Vision Egg: an Open-Source Library for Realtime Visual Stimulus Generation. Front Neuroinformatics 2(4) (2008), doi:10.3389/neuro.11.004.2008
28. NEV 2.0 (Neural Event Format) format specification, http://cyberkineticsinc.com/NEVspc20.pdf
29. NeurALC., http://neuralc.sourceforge.net
30. Mancuso, M., Poiuzzi, R., Rizzotto, G.G.: A fuzzy filter for dynamic range reduction and contrast enhancement. In: Proceedings of the Third IEEE Conference on Fuzzy Systems, IEEE World Congress on Computational Intelligence, vol. 1, pp. 264a, 265–267 (1994)

A Novel Approach for Quantitative Analysis of 3D Phosphenes

C. Soto-Sanchez[1,2], Andrés Olmedo-Payá[1], D. de Santos-Sierra[1], C. Agullo[1], and Eduardo Fernández[1,2]

[1] Bioengineering Institute, University of Miguel Hernandez, Elche, Spain
[2] CIBER-BBN, Zaragoza, Spain

Abstract. Visual prosthesis development relays in the ability of the visual system to evoke visual topographically organized perceptions called phosphenes when it is electrically stimulated. There have been many approaches to quantify phosphenes and describe their position in the visual field but no one managed to be accurate and easy to be handled by visually impairment individuals. Here, we present a highly accurate, intuitive and freely movement method to draw phosphenes in the 3D visual space. We use an infrared sensor from the commercial Kinect hardware through a customized software to detect the movements of the subjects drawing on the air in the same 3D coordinate were they perceive the phosphenes.With this new technique we introduce the component of depth of the visual space in the phosphenes mapping, a disregarded component in the old 2D techniques. No techniques in the past had this measurement in account but our results show that transcraneal magnetic stimulated subjects clearly perceived phosphenes at different depth locations (data not shown at this paper). Furthermore this new mapping technique has three main advantages: (i) it allows patients to locate phosphenes in the real 3D visual space were they perceive the phosphenes, (ii) allows a quantitative measurement of the area and shape of the phosphenes (iii) and avoid the interactions between external devices and patients in order to facilitate the performance to the low vision or blind individuals.

1 Introduction

Development of visual prosthesis pursue a challenging objective to restore the lost visual function to the blind people. To this purpose it is crucial to be able to transform the visual information from the external world into biological signals reproducing the lost visual function. Prosthesis emulate the physiological electrical function of the nervous system generating electrical impulses similar to those of the nervous system working as an interface between the nervous system and the external world. In the visual system the visual perceptions evoked by electrical stimulation are called phosphenes. These visual perceptions are topographically organized representing the visual field and they can be mapped stimulating the visual system in a topographic manner [1–3]. But one of the

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 342–349, 2013.

challenges in this research field it is to match the stimulation places in the brain to the real visual world. No matter the stimulation method used is electrical, magnetic or optogenetic [4, 5] or the visual structure stimulated is the visual cortex [4, 6, 7], the optical nerve [8] or retina [9, 10] the need to map phosphenes always exists. For that reason is necessary to know the stimulation map to match the visual perceptions to their location in the real world. This technique it is called phosphene mapping and it is been used to map phosphenes using different strategies like drawing methods in a 2D board [11, 12] touch screen methods [13] and pointing methods [4, 6]. A main drawback arising from the previous methods is that the subjects have to reinterpret what they perceive and then transform that information in a 2D and limited format like a board or a screen. This artificial transformation reduces the scale of the visual space and the visual perceptions in an intangible manner loosing the possibility of drawing and localizing phosphenes in the visual field accurately. With this new approach the subjects draw directly in the real visual space in the precise 3D coordinate were they perceive the visual sensation. Furthermore and given the difficulty of working with patients having little or no vision, this technique allows the patients to work without external references because the sensor tracks the references at all moments during the entire experiment. Likewise the wireless condition of the whole procedure makes this approach much more convenient than the previous ones for all kind of subjects. The new 3D component we are able to measure it is crucial to develop visual prosthesis in order to be able to navigate through the world around blind individuals.

2 Materials/Methods

2.1 Image Capture

Hardware. We used a Kinect sensor for the Xbox 360 video game system [14] (Microsoft; Redmond, Washington) connected to a PC to track the drawings the subjects performed. Kinect was designed initially for natural interaction in a computer game environment [15] but it is becoming a fashionable tool in many scientific fields [16–18] including biomedicine [19, 20]thank to its 3D data acquisition capture. The Kinect sensor was connected to the USB port of a Intel Core 2 Duo 2.8GHz processor desktop computer with 2GB RAM, running Windows7 ultimate 32bits operating system. Also supported by Windows XP (service pack3). The Kinect is a depth camera consisting of an infrared laser projector (IR Projector, OG12/0956/D306/JG05A) emitting a beam of speckles and an infrared camera (IR CMOS, Microsoft/X853750001/VCA379C7130) receiving the speckles deviation caused by the objects in front of it. The Kinect measures distance by comparing the deviation of the speckles caused by the objects to a reference depth pattern creating a depth image. The operation range of the depth sensor is 0.8-3.5m and the x/y/z spatial resolution at 2m distance is 3/3/10mm [15]. We placed the sensor over a range of 1.5 to 2.8 meters away to the reference points to take the measurements. It was placed aligned with the patients. The Kinect was mounted on top of a 3D plane height adjustable tripod.

Software. We used the OpenNI (Open Natural Interaction) software platform released by Primesense 2010 (Tel-Av iv, Israel). The OpenNI platform includes the SensorKinect091-Bin-Win32-v5.1.0.25 drivers to communicate with the infrared sensor. Data from the Kinect was collected using Kinect NiUserTracker function (OpenNI-win32-1.5.2.23 [21]). We used the nite-win32-1.5.2.7 libraries to recognize the human body segments. The human body segments recognized are head, neck, torso center, shoulder, elbow, hands, hips, knees and feet.

Visual Field Center Selection. We used the nite-win32-1.5.2.7 libraries to recognize the head segment and select the head node located at the middle of the head between the eyes. We used it as the reference point to calculate the center of the visual field. We tracked this point continuously during the trials at the same time as tracking the phosphenes in order to get a reliable reference for the visual field center. We called this point head reference point (HRP).

Pointer Detection. We used the nite-win32-1.5.2.7 libraries to recognize the hand segment and select the hand node. Then we developed a function we called *get_fingertip_tracking_point* to find the closest point of the hand to the Kinect. We calculated the closest point analyzing a space of a 15cm radius circle centered at the hand node and calculating the closest point to the Kinect sensor inside the circle (Fig. 1). We called this point fingertip tracking point (FTTP).

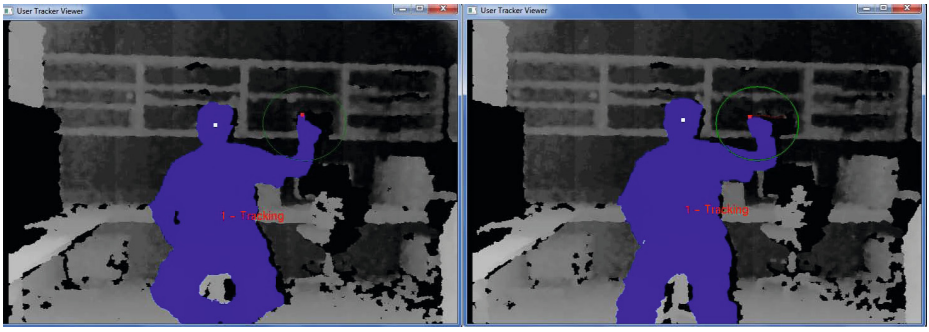


Fig. 1. Left, image capture showing the 15 cm radius circle (green) to calculate the closest data point (red) to the Kinect sensor. Image capture of the user tracker viewer showing the head reference point marked in white and the fingertip tracking point marked in red. Right, image capture showing the trail of the drawing performed by the subject.

Finding the closest point to the Kinect inside the circle allows to track either a fingertip point or any kind of pointer the person holds in his hand. Figure 1. is an image capture showing the fingertip marked in red and the 15cm radius circle centered at the hand. Holding a pointer in their hand gives to the subjects a reference to focus their attention making the drawing task feel more natural and

reliable. We also improved the *get_fingertip_tracking_point* function to stabilize the flickering of the sensor tracking. For each trial we got the x, y, z data points both for the head reference point and for the fingertip tracking point. Figure 2 right shows an image capture that represent visually the trail corresponding to the data captured by the Kinect tracking a subject drawing. The movements described by the pointer were tracked and saved to a TXT file format.

We also developed a function called *get_gesture* to recognize the actions of start and stop the drawing movements. This function recognizes the gesture of the subjects moving their hand to the front to start the drawing task and sets a initial point to start storing data. It recognizes also when the patient withdraws the hand to stop collecting data. This gesture recognizing function avoids unnecessary data collection.

3D Visor Software. We designed a 3D visor software in C++ with OpenGL to visualized and save the drawings data points in the 3D space and measure the size and area of the visual perceptions. To represent the phosphenes we calculated the distance between the head reference point taken as 0,0,0 reference point and the fingertip tracking point. The 3D visor software allows to display and save the distances data in different units such as millimeters or visual degrees. In the Figure 2 we show two image captures of the 3D visor showing a frontal view of a phosphene drawn by a voluntary stimulated with magnetic resonance. The data are shown in visual degrees. Concentric circles centered at the center of the visual field are represented each 10 visual degrees.

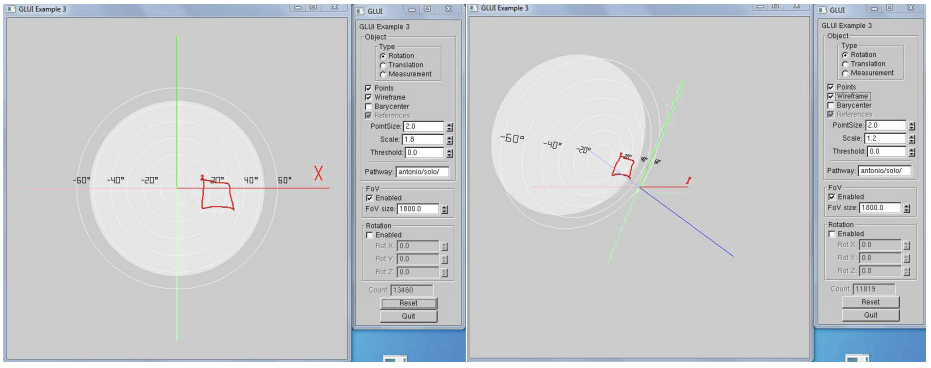


Fig. 2. 3D visor software in C++ with OpenGL. Left, frontal view of a drawing in the 3D visor. Right, rotate view as in the left.

We used the function called *get_gesture* to recognize the actions of start and stop the drawing movements. This function recognizes the gesture of the subjects moving their hand to the front to start the drawing task and sets a initial point to start storing data. It recognizes also when the patient withdraws the hand to stop collecting data. This gesture recognizing function avoids unnecessary data collection.

3 Results

3.1 Software Tracking Accuracy

We tested the error introduced by our software to track the head reference point and the fingertip tracking points. To calibrate our software error we took 5 measurements in each of the 3D axes at 3 different distances inside of the operation range for the Kinect sensor (0.8-3.5m) [15]. We tested the accuracy of the software at 1.5m, 2.25m and 3.00m for the head reference point and 50cm ahead of these positions for the fingertip tracking point corresponding to the natural position of the an adult arm drawing on the air at the front. These measurements were selected based on previous works [22], the operation range of the Kinect and taken in account convenient room dimensions to work with blind patients. To avoid additional movement errors like the human muscular shacking we used a dummy human shaped to take the measurements. The Kinect sensor recognized the dummy as a human body figure tracking the head and fingertip nodes exactly as if it were a human body. The dummy remained in the same fixed position for each of the measurements at each distance. Once the dummy was positioned in a stable position we measured the flickering given by the software for both the head reference and fingertip tracking points. The Kinect system acquired an equal number of samples (200 samples) form each target position. We repeated each of the measurements for five times at each position and calculated the error of our measurement for the x,y,z in common. For that, we define a spatial position as $p_i = (x_i, y_i, z_i)$ where $1 \leq i \leq n$, where n is the number of samples and $\bar{p} = (\bar{x},\bar{y},\bar{z})$ is the media position.

Then we calculate the mean error (ε) as in equation 1,

$$\varepsilon = \bar{e} = \frac{\sum_{i=1}^{n} e\,(p_i, \bar{p})}{n} = \frac{\sum_{i=1}^{n} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2}}{n} \quad (1)$$

where $e(p_i, \bar{p})$ is the euclidean distance of each position to its media. We got five error values, one for each trial giving the range of the error. This error calculation allows to compare the error of the measurements in the three spatial coordinates at the same time. Tables 1a and 1b show the range error results for each of the five measurements at the 3 different distances for the head reference point given in milimeters (Table 1a) and visual degrees (Table 1b). Tables 2a and 2b show the same data as in Table 1a and 1b for the fingertip tracking point.

Table 1a	1.50m	2.25m	3.00m
HRP E1 x,y,z(mm)	1.51	1.94	7.33
HRP E2 x,y,z(mm)	1.50	1.89	7.11
HRP E3 x,y,z(mm)	2.15	1.81	6.22
HRP E4 x,y,z(mm)	1.46	1.70	7.24
HRP E5 x,y,z(mm)	1.14	2.03	6.26

Table 2a	1.00m	1.75m	2.50m
FTTP E1 x,y,z(mm))	2.89	1.64	4.27
FTTP E2 x,y,z(mm)	1.96	2.25	5.48
FTTP E3 x,y,z(mm)	0.78	1.49	9.03
FTTP E4 x,y,z(mm)	2.25	1.45	9.35
FTTP E5 x,y,z(mm)	0.96	2.47	5.29

Table 1b	1.50m	2.25m	3.00m
HRP E1 visual deg	0.06	0.05	0.14
HRP E2 visual deg	0.06	0.05	0.13
HRP E3 visual deg	0.08	0.05	0.12
HRP E4 visual deg	0.05	0.04	0.14
HRP E5 visual deg	0.07	0.05	0.12

Table 2b	1.00m	1.75m	2.50m
FTTP E1 visual deg	0.11	0.04	0.08
FTTP E2 visual deg	0.07	0.06	0.10
FTTP E3 visual deg	0.03	0.04	0.17
FTTP E4 visual deg	0.08	0.04	0.18
FTTP E5 visual deg	0.04	0.06	0.10

The maximum error range for the head reference point at 1.5m distance was 2.15mm and 2.03mm at 2.25m distance but it went up to 7.33mm at 3.00m distance. For the FTTP was 2.89mm at 1.00m distance and 2.47mm at 1.75m distance but it went up to 9.35mm at 2.50m distance. When calculated the error as a function of the visual degrees the maximum error range for the HRP was 0.08 degrees at 1.5m distance and 0.05 visual degrees at 2.25m but it went up to 0.14 visual degrees at 3.00m distance. For the FTTP was 0.11 visual degrees at 1.0m distance and 0.06 visual degrees at 1.75m distance and went up to 0.18 visual degrees at 2.50m distance. The graphs 1 and 1b show the data visually. For all the measurements the minimum error was located at a distance near 2.00m for both HRP and FTTP. Even the error range at 1m (HRP) and 1.5m (FTTP) increased slightly, it was very similar to the error at a distance near 2.00m. The metric error given in mm was quite low and represented a very small part of the visual field (1b and 2b graphs). Even when the error increased drastically at distances of 3.00m for the HRP and 2.5m for the FTTP the visual field equivalent area represented a very small part of the visual field, in all cases less than 0.2 visual degrees.

4 Discussion

Advantages of this new 3D mapping technique are the technical simplicity and the short time required to obtain a phosphene map. It is completely wireless, avoiding the interactions between subjects and external devices in order to facilitate the performance to the low vision or blind individuals. This is especially convenient in the operating room with all the equipment and time restrictions when we deal with visually impairment people. In this paper we described a protocol to systematically map the position and size of phosphenes in the 3D space, and it is compatible both with absolute[4, 23] and relative [24, 25] phosphene mapping procedures. This new technique allows the patients to locate phosphenes in the real 3D visual space where they perceive the phosphenes avoiding the mental abstraction of transforming their perceptions from the real space into a reduced 2D format. Drawing phosphenes in the actual space were they are perceived endows this technique with the accuracy the previous ones lack. It allows quantitative measurement of the area, shape and 3D position of the phosphenes in the actual visual field within an accuracy of milimeters. The error generated by our system is low enough to measure the position and size of phosphenes in the 3D real space. It yields low errors such as 0.05 visual degrees for the HRP and 0.06 visual degrees for the FTTP at the best distance. We consider that the errors of the visual perception and the muscular precision of drawing are much bigger than the system error, so we can disregard the error system if we map phosphenes at the distances around 2 meters. If we think in terms of mapping the visual field to navigate through the environment avoiding obstacles and detecting borders this technique shows a very precise detection and it could be used as a standardized testing system in the framework of the visual prosthesis development, patient testing and for studies regarding the physiological organization of human visual

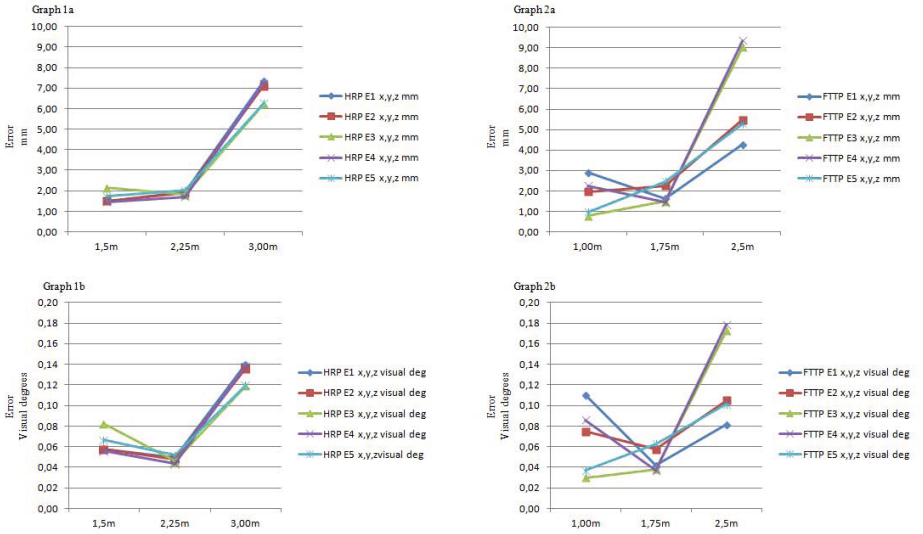


Fig. 3. Top graphs, represent the error intervals for the head reference point (left) and the fingertip tracking point (right) given in mm. Bottom graphs, represent the error intervals in visual degrees; left, head reference point; right, fingertip reference point.

system. This new procedure could be used for standardized testing and help to identify suitable candidates that might benefit from a visual neuroprosthetic device as well as for assessments of device efficacy.

Acknowledgments. We would like to thank Antonio Alarcon for his help collecting data. This work has been supported in part by the CIBER-BBN from the Spanish Government and by the ONCE (National Organization of the Spanish Blind).

References

1. Engel, S.A., Glover, G.H., Wandell, B.A.: Retinotopic organization in human visual cortex and the spatial precision of functional mri. Cereb Cortex 7, 181–192 (1997)
2. Fitzgibbon, T., Taylor, S.F.: Retinotopy of the human retinal nerve fibre layer and optic nerve head. J. Comp. Neurol. 375, 238–251 (1996)
3. Tusa, R.J., Palmer, L.A., Rosenquist, A.C.: The retinotopic organization of area 17 (striate cortex) in the cat. J. Comp. Neurol. 177, 213–235 (1978)
4. Brindley, G.S., Lewin, W.S.: The sensations produced by electrical stimulation of the visual cortex. J. Physiol. 196, 479–493 (1968)
5. Marg, E., Rudiak, D.: Phosphenes induced by magnetic stimulation over the occipital brain: description and probable site of stimulation. Optom. Vis. Sci. 71, 301–311 (1994)

6. Gothe, J., Brandt, S.A., Irlbacher, K., Roricht, S., Sabel, B.A., Meyer, B.-U.: Changes in visual cortex excitability in blind subjects as demonstrated by transcranial magnetic stimulation. Brain 125, 479–490 (2002)
7. Schmidt, E.M., Bak, M.J., Hambrecht, F.T., Kufta, C.V., O'Rourke, D.K., Vallabhanath, P.: Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. Brain 119(pt. 2), 507–522 (1996)
8. Veraart, C., Wanet-Defalque, M.-C., Gerard, B., Vanlierde, A., Delbeke, J.: Pattern recognition with the optic nerve visual prosthesis. Artif Organs 27, 996–1004 (2003)
9. Humayun, M.S., Weiland, J.D., Fujii, G.Y., Greenberg, R., Williamson, R., Little, J., Mech, B., Cimmarusti, V., Van Boemel, G., Dagnelie, G., de Juan, E.: Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. Vision Res. 43, 2573–2581 (2003)
10. Rizzo III, J.F., Wyatt, J., Loewenstein, J., Kelly, S., Shire, D.: Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. Invest Ophthalmol Vis. Sci. 44, 5355–5361 (2003)
11. Cowey, A., Walsh, V.: Magnetically induced phosphenes in sighted, blind and blindsighted observers. Neuroreport 11, 3269–3273 (2000)
12. Fernandez, E., Alfaro, A., Tormos, J.M., Climent, R., Martinez, M., Vilanova, H., Walsh, V., Pascual-Leone, A.: Mapping of the human visual cortex using image-guided transcranial magnetic stimulation. Brain Res. Brain Res. Protoc. 10, 115–124 (2002)
13. Zhang, L., Chai, X., Ling, S., Fan, J., Yang, K., Ren, Q.: Dispersion and accuracy of simulated phosphene positioning using tactile board. Artif Organs 33, 1109–1116 (2009)
14. Microsoft, Kinect (2010), `http://www.xbox.com/en-us/kinect/`
15. PrimeSense (2010), `http://www.primesense.com/`
16. Andrew, D.: Using a depth camera as a touch sensor. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, pp. 69–72 (2010)
17. Benavidez, M., Jamshidi, P.: Mobile robot navigation and target tracking system. In: Proceedings of the 6th International Conference on System of Systems Engineering: SoSE in Cloud Computing, Smart Grid, and Cyber Security, pp. 299–304 (2011)
18. Bainbridge-Smith, A., Stowers, J., Hayes, M.: Altitude control of a quadrotor helicopter using depth map from microsoft kinect sensor. In: Proceedings of the IEEE International Conference on Mechatronics, pp. 358–362 (2011)
19. Bo, A.P.L., Hayashibe, M., Poignet, P.: Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 2011, pp. 3479–3483 (2011)
20. Chang, Y.-J., Chen, S.-F., Huang, J.-D.: A kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res. Dev. Disabil. 32(6), 2566–2570 (2011)
21. OpenNI (2011), `https://github.com/OpenNI/OpenNI`
22. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors (Basel) 12(2), 1437–1454 (2012)
23. Dobelle, W.H., Mladejovsky, M.G.: Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. J. Physiol. 243, 553–576 (1974)
24. Dobelle, W.H.: Artificial vision for the blind by connecting a television camera to the visual cortex. ASAIO J. 46(1), 3–9 (2000)
25. Everitt, B.S., Rushton, D.N.: A method for plotting the optimum positions of an array of cortical electrical phosphenes. Biometrics 34, 399–410 (1978)

Inverse Allometry: Foundations for a Bioinspired LVH-Prediction Model

María Paula Bonomini[1], Fernando Ingallina[2], Valeria Barone[2], and Pedro D. Arini[1,3]

[1] Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina
[2] Instituto de Investigaciones Médicas Dr. A. Lanari, Universidad de Buenos Aires (UBA)
[3] Instituto Argentino de Matemática, CONICET, Buenos Aires, Argentina
pbonomini@fi.uba.ar

Abstract. Allometry, in general biology, measures the relative growth of a part in relation to the whole living organism. Left ventricular hypertrophy (LVH) is the heart adaptation to excessive load (systolic or diastolic) that leads to an increase in left ventricular mass, which in turn, raises the electrocardiographic voltages. If this mass increase followed an allometric law, then it would be possible to design a bioinspired model based on the allometric equation to predict LVH. In this work, we first investigated the validity of this hypothesis and then proposed an LVH marker based on the inverse allometry model. Based on clinical data, we compared the allometric behavior of three different ECG markers of LVH. To do this, the allometric fit AECG $= \delta + \beta$(VM) relating left ventricular mass (estimated from echocardiographic data) and ECG amplitudes (expressed as the Cornell-Voltage, Sokolow and the ECG overall voltage indexes) were compared. Besides, sensitivity and specificity for each index were analyzed. The more sensitive the ECG criteria, the better the allometric fit. Finally, the Receiver Characteristic Curve (ROC) of an allometric model proposed here was computed. In conclusion: The allometric paradigm should be regarded as the way to design new and more sensitive ECG-based LVH markers.

1 Introduction

Left ventricular hypertrophy (LVH) is the heart way to adapt to overloads, either during diastolic or systolic periods. This adaptation consists of increasing the diameter of the cardiac fibers and, consecuently, of left ventricular mass. Such augmented mass directly affects the electrocardiographic signal by raising its voltage amplitude. We address two questions here. Firstly: does this increase in amplitude keep an allometric relationship with the increase in left ventricular (LV) mass? To figure this out, a comparison of the allometric adjustment of different LVH indexes was carried out. Secondly: can the allometric model be the foundation for more sensitive bioinspired LVH markers?

Allometry, in general biology and physiology, has been recently reassessed with some new insights as, for example, the relationship between organ and animal size [1], organ and biological times such as heart rate, volumes and capacities

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 350–357, 2013.

such as cardiac output [2], ECG-PR interval and animal size [3] and finally, electrocardiographic parameters in ischemia and number of cardiac diseased fibres [5, 4].

The existence of many different criteria for diagnosing LVH makes clinical application more complex. The sensitivity of the LVH indexes based on ECG is generally quite low (usually less than 50%), while the specificity is quite high (often in the range of 85% to 90%) [6][7][8]. Published studies are currently insufficient to indicate whether any of the more recently proposed criteria are clearly superior to the others or are simply redundant. For these reasons, the aim of this work is to provide insight into new clues and theoretical support to help researchers find more sensitive markers of LVH.

2 Materials and Methods

2.1 Patient Population

According to the Penn Convention and using the Deveraux equation [14], LV mass was assessed in 36 patients, from which, 23 out of 36 showed echocardiographic mass index greater than 259 g and 166 g for men and women, respectively, leading to LVH diagnosis. The average age of the studied population was composed of 17 men (75.44±8.13 years old) and 19 women (72.75±13.54 years old). Pathologies varied, since recruitment was done on outpatients of a general hospital. Patients with intraventricular conduction diseases (IVCD) were ruled out, since both LVH and IVCDs alter QRS patterns, therefore, the existence of an IVCD may impact the accuracy of ECG criteria for LVH [10].

2.2 ECG Criteria for LVH

We compared the allometric adjustment of three ECG markers for LVH, i.e., Cornell index, ECG total 12-lead voltage and Sokolow index. They are widely used in clinical practice. These indexes were calculated as follows:

- Cornell (voltage) index: There are two versions of this index, one concerning voltage only and the other one combining QRS voltage and duration. We used here only the voltage version that combines the amplitude of the S wave in V3 lead and the amplitude of R wave in aVL lead [10, 11].
  - Men: $S_{V3} + R_{aVL}$ >2,8 mV (28 mm)
  - Women: $S_{V3} + R_{aVL}$ >2,0 mV (20 mm)
- Total 12-lead voltage: Total 12-lead voltage, measured as the sum of all S and R peaks of all 12 leads > 175 mm (8, 10).
- Sokolow index: who in 1949 introduced the widely used criterion based on the sum of the amplitude of the S wave in V1 lead plus the amplitude of R wave in V5 or V6 leads. The cut-off point for this index is >/= 3,5 mV (35 mm) (8, 11). We have chosen V6 lead for our analysis.
  - $S_{V1}+R_{V6}$ >= 3,5 mV (35 mm)

For the ECG studies, we have used a standard 12-lead ECG device and obtained 10 second-recordings. The device had a sampling rate of 400 Hz and 12 bits resolution. From the ECG recordings, peak amplitudes were averaged out from all the beats contained in the entire recording.

2.3 Left Ventricular Mass Calculation

In the late 80s, Levy and coworkers published a landmark paper evaluating a subset of individuals without known cardiovascular risk factors in the Framingham Cohort [10]. These authors calculated LV mass both with the ASE (American Society of Echocardiography) convention and Troy equation [13]:

$$LV_m(Troy) = 1.05 * ([LVID_d + PWT_D + IVST_d]^3 - LVID_d^3) \quad (1)$$

Where: $LVID_d$ = Left Ventricular Internal Diameter in Diastole PWT_d = Posterior Wall Thickness in Diastole $IVST_d$ = Interventricular Septum Thickness in Diastole and with the Penn Convention and Devereux equation [15]:

$$LV_m(Dev) = 1.04 * ([LVID_d + PWT_d + IVST_d]^3 - LVID_d^3) - 13,6g \quad (2)$$

In this work, the Penn Convention and Devereaux equation were chosen. The authors proposed normal limits for LV mass for men and women, based on cut points at two standard deviations above the mean [9, 15].

2.4 Allometric Equation

The term allometry was first used by Snell, in 1891 [16], to express the mass of a mammal's brain as a function of the body mass. The growth velocity of a component, y, is related to the growth velocity of another component (or the whole organism), x, in a constant way. This was clearly described by von Bertalanffy in 1957 [17]. Thus, the relative rate of change of a given event, y, is proportional to the relative rate of change of body mass or body weight, x, i.e.,

$$(dy/dt)/y = B(dx/dt)/x \quad (3)$$

After integration and some easy algebraic manipulation, equation (3) becomes

$$y = Ax^B \quad (4)$$

The parameters A and B require numerical estimation by an appropriate procedure usually using empirical information. By the same token, let us say that the amplitude of the ECG (we use Aecg in a general form, since Aecg will be quantified as the ECG criteria for LVH) follows a relationship with the number of ventricular hypertrophic fibres, and therefore, the ventricular mass (Vm). Thus, equation 4 can be reformulated as,

$$Aecg = \alpha * Vm^\beta \quad (5)$$

After taking logarithms of both sides, the latter equation becomes

$$ln(Aecg) = ln(\alpha + \beta * ln(\gamma)) + \beta * ln(Vm) \tag{6}$$

which can be reduced to

$$AECG = \delta + \beta(VM) \tag{7}$$

We define AECG as ECG voltage or amplitude, where $\delta = \ln(\alpha) + \beta\ln(\gamma)$, VM = ln(Vm) and AECG = ln(Aecg). The straight line, equation 5, in log-log plot with the parameters β and δ would represent the increase in ECG amplitude as function of the amount of hypertrophic fibres.

2.5 Numerical Procedure

To calculate the two constants δ and β and later on apply the mathematical expression in equation 7, linear regression was implemented in order to evaluate the allometric fit on a log-log plot gathering the variables ventricular mass and LVH index. Notice that δ represents the intercept to the origin and β is the slope of the regression line when plotted on log-log coordinates. When β ¡1, the relative weight of the ECG criteria is greater in small than large hypertrophy. Also, sensitivity and specificity was estimated from our set of data by calculating the following equations: $Sensitivity = TP/(TP + FN)$ and $Specificity = TN/(TN + FP)$, where TP are the true positive, FP the false positive and TN the true negative cases, all of them confirmed by echocardiographic analysis. More specifically, echocardiographic mass calculation, as described above, and the criteria of a cardiologist were set as the gold standard for LVH diagnosis.

2.6 ROC Curves

We set δ=1 and β=1 and calculated the Receiver Operating Characteristic (ROC) curves for the simplified allometric model log(y) = log(m) and plotted the Sensitivity against the 1-Specificity values for the different possible cut-off points. Thereafter, the optimal cut-off point in the ROC curve was computed as the point nearest the top left-hand corner. This selection maximizes the sensitivity and specificity sum, when it is assumed that the 'cost' of a false negative result is the same as that of a false positive result [18].

3 Results

3.1 Allometry and Sensitivity of Electrocardiographic LVH Markers

Using the equation AECG = δ + β (VM) in log-log representation, the fitting procedure produced coefficients for every LVH index. Moreover, graphic results for the allometric adjustments on log-log plots are showed in Figure 1. Linear

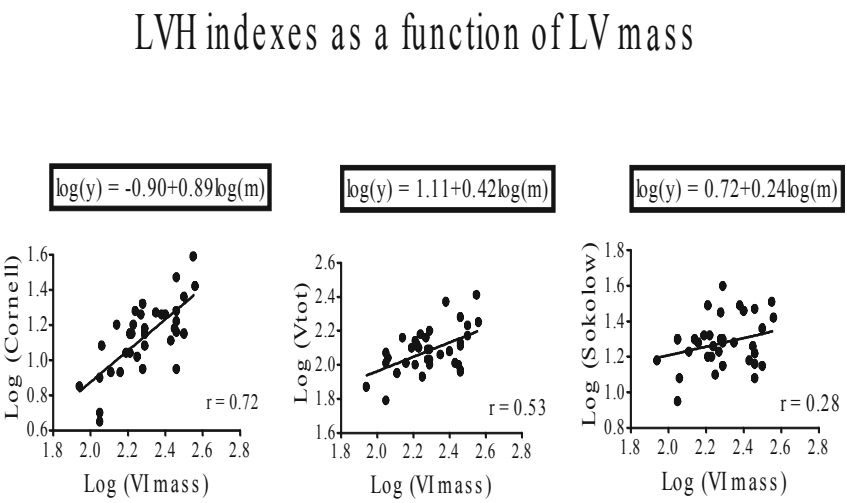


Fig. 1. Linear regressions for the allometric fits of ECG-based LVH markers: Cornell index (left panel), total 12-lead voltage (middle panel) index and Sokolow index (right panel).

Table 1. Mean ± SD of SVD indexes for healthy subjects (PTB Database) and ischemic patients (STAFF-III Database), $p<0.005$

	Sensitivity	Specificity	δ	β	r
Voltage-Cornell	42%	96%	-0.90	0.89	0.72
Sokolow	22%	100%	1.11	0.42	0.53
Total Voltage	17.2%	95%	0.72	0.24	0.28

regression and determination coefficients r are also displayed. The best correlation was offered by the Voltage-Cornell index (r=0.7229), followed by the Total Voltage index (r=0.5292) and last, by Sokolows (r=0.2769), the latter obviously very low.

Notice as well, that all the slopes β in the adjustment are positive and smaller than 1. This means that indexes will increase along with the amount of hypertrophy up to a certain saturation level. This is, there will be a point at which even though the amount of hypertrophic fibres increases, the LVH index will remain almost constant. Slopes greater or even equal to 1 would be more desirable

in order not to lose sensitivity with the hypertrophy extent. Table 1 shows the sensitivity and specificity as reviewed in the bibliography and as estimated from our set of data. In all cases, notice how the sensitivity suffered when calculated from a reduced set of samples.

3.2 Theoretical Allometric Model

Finally, the theoretical allometric model was computed on the left ventricular mass data available by setting the coefficient $\delta = 0$ and $\beta = 1$. Afterwards, the ROC curve of this theoretical model and the remaining ECG-based indexes were calculated in order to compare their hability to predict LVH. Also, the area under

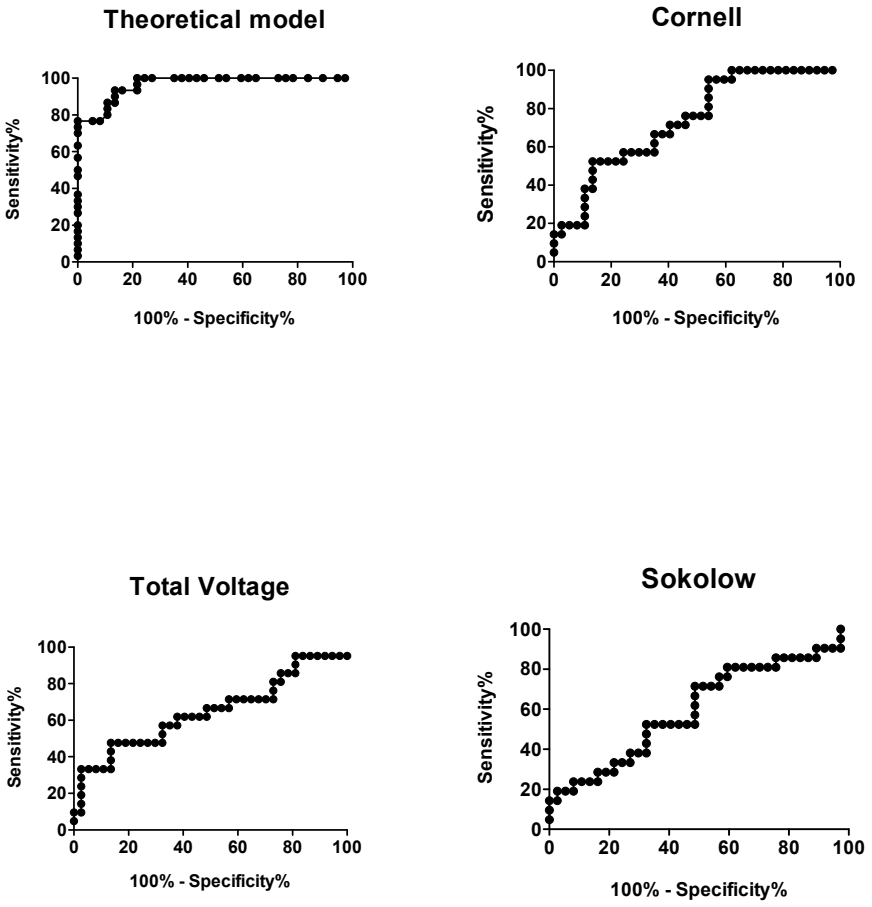


Fig. 2. ROC curve for the theoretical allometric model (upper left panel, AUC=0.96), the Cornell index (upper right panel, AUC=0.74), Total Voltage index (lower left panel, AUC=0.64) and Sokolow index (AUC=0.59)

the curves (AUC) were calculated. Figure 2 shows the great potential of the allometric model to predict LVH, displaying an AUC = 0.96, clearly outstanding from the others. AUC values for the LVH indexes were as follows: Theoretical model (AUC=0.96), Cornell (AUC=0.74), Total Voltage (AUC=0.64) and Sokolow (AUC=0.59).

4 Discussion and Conclusions

The results show that sensitivity goes along with allometric behavior when searching for LVH markers based on ECG. However, certain constraints should be regarded. For instance, the population under study was quite homogeneous in terms of age, presenting mainly the same type of hypertrophy. Thus, the analysis herein accomplished holds for the elderly only. It is important to notice that patients with intraventricular conduction diseases (IVCD) were excluded from the study in order not to confuse the symptoms, since both LVH and IVCD can lead to similar changes in QRS. Another limitation of the study is that not all the QRS-based indexes of LVH were analyzed. Nevertheless, the results found here encourage a more complete study including all electrocardiographic indexes of LVH over a larger and more heterogeneous sample population. Interestingly, the Voltage-Cornell index showed the best allometric fit, with a sensitivity of 42%, as collected from the bibliography [10] and 35%, as estimated from our data pool. One point to mention is the slope β of the linear regressions in the allometric fit. All of them resulted positive and smaller than 1, which means that the greater the amount of hypertrophic fibres, the lesser the growth of the LVH marker, leading to a loss of sensitivity with hypertrophy extent. Therefore, if more sensitive markers are required, β tending to 1 would be desired. It is important to note that even though the latter index was designed for different LVH cut-off points (132 g/m2 for men and 109 g/m2 for women) as used here, it adapted well to the new definition, postulating itself as the most robust ECG marker of hypertrophy. Finally, the voltage increase, as expressed in the electrocardiographic indexes studied here followed an allometric relation with left ventricular mass. This fact should guide the search of new and more sensitive markers of hypertrophy by taking into account the allometric law. Moreover, slopes of the regression lines (or constant β) resulted all positive and smaller than one, leading to ECG-based indexes showing a saturation for high hypertrophy levels. To overcome this, a theoretical model with $\beta = 1$ was proposed, and an AUC of 0.96 was found, proving the power of the allometric model to predict LVH.

Acknowledgments. To the General Hospital Instituto de Investigaciones Médicas Dr. A. Lanari, belonging to University of Buenos Aires (UBA). Besides, the first author (MPB) and PDA are career investigators of CONICET.

References

[1] Lindstedt, S.L., Schaeffer, P.: Use of allometry in predicting anatomical and physiological parameters of mammals. Laboratory Animals 36, 1–19 (2002)

[2] Lindstedt, S.L., Miller, B., Buskirk, S.: Home range, time, and body size in mammals. Ecology 67(2), 413–418 (1986)

[3] Noujaim, S.F., Lucca, E., Muñoz, V., Persaud, D., Berenfeld, O., Meijler, F.L., Jalife, J.: From mouse to whale: A universal scaling relation for the PR interval of the electrocardiogram of mammals. Circulation 110, 2801–2808 (2004)

[4] Bonomini, M.P., Arini, P.D., Gonzalez, G., Buchholz, B., Valentinuzzi, M.E.: The allometric model in chronic myocardial infarction. Theoretical Biology and Medical modeling, epub. 9, 15 (2012)

[5] Bonomini, M.P., Arini, P.D., Valentinuzzi, M.E.: Probability of ventricular fibrillation: allometric model based on the ST deviation. Biomed. Eng. Online 10(2) (January 13, 2011)

[6] MacFarlane, P.W., Lawrie, T.D. (eds.): Comprehensive Electrocardiography: Theory and Practice in Health and Disease. Pergamon Press, Oxford (1998)

[7] Hsieh, B.P., Pham, M.X., Froelicher, V.F.: Prognostic value of electrocardiographic criteria for left ventricular hypertrophy. Am Heart J. 150(1), 161–167 (2005)

[8] Hancock, W., Deal, B., Mirvis, D., Okin, P., Kligfield, P., Gettes, L.: Recommendations for the standardization and interpretation of the electrocardiogram. Journal of the American College 53(11) (2009)

[9] Casale, P., Devereux, R., Kligfield, P., Eisenberg, R.R., Miller, D.H., Chaudhary, B.S., Phillips, M.C.: Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. J. Am. Coll. Cardiol. 6, 572–580 (1985)

[10] Levy, D., Savage, D., Garrison, R.J., Anderson, K.M., Kannel, W.B., Castelli, W.P.: Echocardiographic criteria for left ventricular hypertrophy: the Framingham Heart Study. Am J. Cardiol. 59, 956–960 (1987)

[11] Siegel, R.J., Roberts, R.W.: Electrocardiographic observations in severe aortic valve stenosis: correlative necropsy study to clinical, hemodynamic, and ECG variables demonstrating relation to 12-lead QRS amplitude to peak systolic transaortic pressure gradient. Am. Heart J. 103(2), 210–221 (1982)

[12] Sokolow, M., Lyon, T.P.: The ventricular complex in left ventricular hypertrophy as obtained by unipolar and limb leads. Am. Heart J. 38(2), 273–294 (1949)

[13] Troy, B.L., Pombo, J., Rackley, C.E.: Measurement of left ventricular wall thickness and mass by echocardiography. Circulation 45(3), 602–611 (1972)

[14] Devereux, R.B., Reichek, N.: Echocardiographic determination of left ventricular mass in man. Anatomic Validation of the Method Circulation 55(4), 613–618 (1977)

[15] Foppa, M., Duncan, B., Rohde, L.: Echocardiography-based left ventricular mass estimation. How should we define hypertrophy? Cardiovasc Ultrasound 3, 3–17 (2005)

[16] Snell, O.: Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen Fähigkeiten (Dependence of brain weight on body weight and the intellectual capacity). Arch. Psychiatr. u Nervenkr. 23, 436–446 (1891)

[17] Von Bertalanffy, L.: Quantitative laws in metabolism and growth. Quaterly Review Biology 32(3) (1957)

[18] Altman, D.: Statistics in medical journals: developments in the 1980s. Stat. Med. 10(12), 1897–1913 (1991)

Cochlear Implant: Transcutaneous Transmission Link with OFDM

Vicente Garcerán-Hernández[1] and Ernesto A. Martínez–Rams[2]

[1] Universidad Politécnica de Cartagena, Antiguo Cuartel de Antiguones (Campus de la Muralla), Cartagena 30202, Murcia, España
vicente.garceran@upct.es
[2] Universidad de Oriente, Avenida de la América s/n, Santiago de Cuba, Cuba
eamr@fie.uo.edu.cu

Abstract. This paper presents the use of the OFDM (Orthogonal Frequency Division Multiplexing) modulation technique in the design and simulation of a telemetry system for a cochlear implant. Data were processed through the Simulink module of MATLAB^{TM}. High transmission speed and high spectral efficiency were achieved with this design and simulation. Two types of OFDM were studied and compared regarding spectral efficiency and noise immunity, and the superiority of OFDM modulation using the QAM (Quadrature Amplitude Modulation) method was shown over OFDM, which uses the DQPSK (Differential Quadrature Phase Shift Keying) modulation method.

1 Introduction

Cochlear implants are devices, a component of which is surgically inserted in the internal ear, partially restoring audition to people with deep deafness. The cochlear implant block diagram is shown in Figure 1.

Sound is first picked up by the microphone and processed by the external processor of sound. The digital sound processor (DSP) is used to analyze the sound and encode results into instruction trains. The DSP also contains memory units or "maps that store patient specific information. The maps and other speech processing parameters can be set or modified by a PC fitting program. The external transmitter/receiver modulates the trains and sends both data and energy to the implant circuit. The internal implant receiver/transmitter recovers data and energy from the RF (Radiofrequency) signals, decodes the instructions and accordingly directs the stimulator to generate electrical stimuli. Finally, the stimuli are applied to electrodes array to excite the auditory nerve [1].

1.1 Transmission Link

A percutaneous connector or transcutaneous link is used to convey stimuli or stimulus information from the external speech processor to the implanted electrodes [2,3]. The percutaneous system transmits the stimuli to the electrodes directly through plug. In this case there is no implanted electronics other than

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 358–367, 2013.

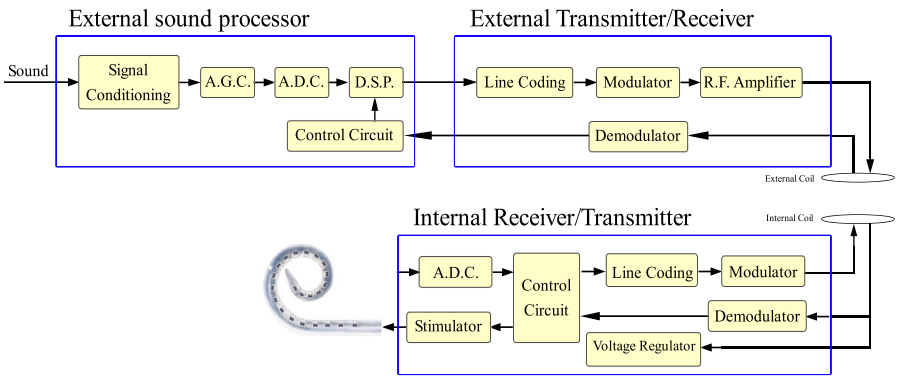


Fig. 1. Block diagram of the cochlear implant system with telemetry. It has four main parts: external controller, external transmitter/receiver, internal receiver/transmitter (implantable unit) and transmission link.

the electrodes. The major advantage of this system is flexibility and signal transparency (the specification of stimuli is no way constrained by the limitations imposed with any practical design of a transcutaneous transmission link). The inconvenience of percutaneous connector was the infection that it could engender in the ear. In transcutaneous link, an external transmitter encodes the stimulus information for RF transmission from an external antenna to an implanted antenna [4]. The internal receiver decodes the signal and delivers the stimuli current to the electrodes. The transmitter and the implanted receiver are held in place by a magnet. The advantage of this system is that the skin is closed over the implant components after the operation, which may reduce the risk of infection. The limitation of this type is that the implanted electronics may fail and would require surgery for replacement. A disadvantage is that only a limited amount of information can be transmitted across the skin. Another limitation of the transcutaneous connector is that it contains magnetic materials, which are incompatible with MRI (Magnetic Resonance Imaging) scanners. Most of the cochlear implant devices today utilize this kind of transmission. In some cases, the link may be bidirectional [5]. The data sent from the external device to the implant, are known as "downlink transmission" or "forward telemetry" and data from the implanted component out to the external antenna and external controller, are known as "uplink transmission" or "reverse telemetry". Such data can include information about the status of receiver/stimulator, impedance of the implanted electrodes, voltages at unstimulated electrodes and intracochlear evoked potentials [5]. Such measurements can be useful in assessing the condition of the auditory nerve and for programming the speech processor [2].

1.2 Telemetry

Bidirectional communication is, most commonly, performed via a wireless, transcutaneous inductive link [6,7]. This battery-less technique minimizes the size of the implant and eliminates patient discomfort. The inductive link channel consists of two closely-spaced, mutually-coupled coils, one implanted and one placed outside the human body. The external unit telemeters power and modulated data (commands and stimulation parameters) to the implant (downlink transmission). It is noted that wireless communication could only deliver power in the range of a few milliwatts [8]. Since a data signal serves the dual purpose of delivering power along with data control signal, the nature of data communication protocol also becomes crucial for power savings. Depending on the modulation format, several values of data transmission rates, error rates, delivered power and circuit complexity can be achieved.

In [9] introduced the design of the speech processor and the prototype implant that sequentially stimulate an array of 8 electrodes (at 4 kHz per electrode and with biphasic pulses) to emulate and test the different cochlear stimulation strategies as yet known, as well as any another variant intended to emulate them in real time. The digital sound processor serially transmits stimuli codes and synchronous codes to the implant. It is necessary to implement a telemetry system (transmitter/receiver) able to intercommunicate both devices keeping communication speed within a small bandwidth. The aim of our research was a telemetry system design with high spectral efficiency and high immunity to noise and interference, and with a control system capable of decoding and cochlear stimulation codes and of addressing at their corresponding electrodes within the electrodes array.

2 Design and Simulation of the Transmission Link

The transmission link for forward power and data telemetry is implemented using two closely-spaced, inductively-coupled coils. In cochlear implants, high data rate transmission toward the implant is required [8]. However, due to the fact that the power dissipation in the tissue increases with the carrier frequency squared [8], this high data rate cannot be achieved by increasing the carrier frequency. Hence, a signal modulation scheme should be chosen to help establish this high data transfer rate link. Between different types of modulation techniques, digital modulation schemes including amplitude shift keying (ASK) [10], binary amplitude shift keying (BASK) [11], binary frequency shift keying (BFSK) [8,12], phase shift keying (PSK) [13,14], binary phase shift keying (BPSK) [15], and quadrature phase shift keying (QPSK) [16] are modulation methods mostly used in implantable device. ASK has no constant amplitude symbols, so reducing the maximum amount of transferred power via the inductive link [6], and achieve low data transmission rates [17]. The carrier using either FSK or PSK has constant amplitude, which may increase the maximum amount of transferred power [16]. FSK and PSK have wider usage than ASK because of their lower sensitivity to amplitude noise. PSK modulation with constant amplitude symbols and fixed

carrier frequency can be the best choice for high-data-rate applications [18,19]. In PSK modulation, bit rates as high as 1.12 Mbps for a carrier frequency of 13.56 MHz have been reported in literature [17]. Unfortunately, RF telemetry has some major disadvantages with respect to achieving a sufficient BW and SNR for operation at high data rates [20]. Since the bandwidth is limited, RF telemetry links often take advantage of the Nyquist relationship, Equation 1, and utilize multiple encoding states.

$$C = 2BWlog_2(M) \tag{1}$$

where M is the number of possible signal states per pulse [21].

ASK, FSK and PSK modulations required high frequency and bandwidth. These modulations can be combined to transmit multiple bits per symbol in order to reduce bandwidth, thus giving rise to QAM (Quadrature Amplitude Modulation) and QPSK. Furthermore, these modulation schemas are linear and required linear amplifiers. It is possible to enhance the spectral performance with M-ary PSK modulation, but it is very susceptible to noise because the low spacing between constellation points. M-ary QAM presents big spacing between constellation points, but it is required very linear amplifiers.

A higher SNR is required for these systems that utilize a large number of signal states than for those using a simple binary encoding scheme. This often translates into high power consumption for large bit rates. RF systems also have disadvantages with respect to achieving a SNR sufficient for operation. The SNR for a radio telemetry link is limited by a number of factors including the transmitter power, inherent electrical noise sources, tissue absorption of the carrier, interfering noise from other RF sources (e.g. wireless network users, wireless telephones, etc.) and the location and separation of the transmitter and receiver. Absorption of the carrier signal by the tissue may also be a significant contributing factor, and increases with the square of the frequency, resulting in a decreased signal to noise ratio for a given power budget when higher frequencies are used (the higher BW bands are at higher frequencies). While transcutaneous RF data transmission that meets the 40 Mbps specification is technically possible, it would be quite difficult to achieve a practical system because of the reasons described above [20]. In information theory, the Shannon-Hartley theorem tells the maximum rate at which information can be transmitted over a communications channel of a specified bandwidth in the presence of noise. Considering all possible multi-level and multi-phase encoding techniques, the Shannon-Hartley theorem states the channel capacity C, meaning the theoretical tightest upper bound on the information rate (excluding error correcting codes) of clean (or arbitrarily low bit error rate) data that can be sent with a given average signal power S through an analog communication channel subject to additive white Gaussian noise of power N, is Equation 2:

$$C = Blog_2(1 + \frac{S}{N}) \tag{2}$$

where C is the channel capacity in bits per second; BW is the bandwidth of the channel in hertz (passband bandwidth in case of a modulated signal); S is the average received signal power over the bandwidth (in case of a modulated signal, often denoted C, i.e. modulated carrier), measured in watts (or volts squared); N is the average noise or interference power over the bandwidth, measured in watts (or volts squared); and S/N is the signal-to-noise ratio (SNR) or the carrier-to-noise ratio (CNR) of the communication signal to the Gaussian noise interference expressed as a linear power ratio (not as logarithmic decibels).

In [9], serial communication speed between the speech processor and the prototype implant is about 1.92 Mbps (without signal modulation), calculated as the product of stimulation frequency per channel (4 kHz) times the number of samples needed to codify each biphasic pulse, as represented in Figure 2 (3 samples) times the number of bits to transmit per sample (20) and times the quantity of electrodes to stimulate (8).

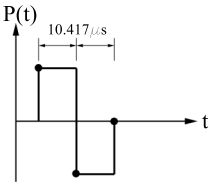


Fig. 2. Form of biphasic stimulation pulse. The circles correspond with the values of the samples that codify each pulse.

For optimum processing of speech signals in real time, sample-block processing was preferred to sample-by-sample processing, since a sample-block is a fixed amount of samples, which makes real time processing possible. Sample-block duration should not be longer than 4 ms in order to achieve visual synchronicity between the movement of the speakers lips and the signals that the implanted person perceives. We chose blocks of 125 μs, yielding 240 bits per block as a result, which is the first requirement of the telemetry system design.

Other design requirements were: that the system should be immune to noise, that it could be adapted to and implemented in an FPGA, because of which the algorithm to develope should allow for exportability towards any of the different FPGAs developed by the Xilinx Company (Spartan-6; Virtex-6). It should at the same time be capable of supporting continuous data-flow systems. It is important for the design of such systems that their control is kept as simple as possible. For this reason, the MATLAB-Simulink environment was selected, since its System Generator lets control be carried out through a simple protocol, which is based on controlling a bit (data valid) that informs the continuous block of when data are available.

To achieve robustness of the system in the face of channel noise, it is necessary to use error detection and error correction mechanisms in spite of the increase in

transmission speed (signal without modulation) caused. In the present application, the FEC (Forward Error Correction) [22] method was preferred to the ARQ (Automatic Repeat reQuest) [23] method. The reason is that this telemetry system must work in real time without delay. During a conversation between people, the delay creates an asynchrony between what is being heard and the speakers lips movement. Within the FEC method, convolutional codification with codification rate equal to 1/2 was chosen before others of greater order, because a higher signal immunity to channel noise was not considered necessary, given the proximity between the transmitter and receiver antennas (just a few millimeters apart; in this case, the thickness of the skin plus a thin layer of mastoid bone).

QAM and QPSK techniques combined with Orthogonal Frequencies Division Multiplexing (OFDM) [24,25] allows a much more efficient use of the transmission link, because OFDM make it possible distribute the information in subchannels or subcarriers. Furthermore, OFDM is also effective when using mutually orthogonal subcarriers which avoids intersymbol interference, an essential property for information recovery. To maintain a desirable characteristics for high-bit-rate transmission with lower bandwidth, we have chosen the OFDM technique.

OFDM effectively partitions the overall system bandwidth into a number of N orthogonal subbands or subchannels. Each of them is associated with a respective subcarrier upon which information may be modulated. For each subcarrier we used digital modulation which are spectrally more efficient as 16-QAM and DQPSK. The latter, additionally simplifies the receiver design because eliminates the phase ambiguity present in the recovered carrier.

2.1 OFDM with DQPSK

The transmitter-receiver scheme using OFDM with DQPSK, simulated through Simulink of MATLAB, is shown in Figure 3. IFFT block is the inverse transform and FFT block is the forward transform. The data generator used generates a bit stream. It is processed using DQPSK modulator to map the input data into symbols. These symbols are now sent through IFFT block to perform IFFT operation of each row of input vector u to generate N parallel data streams. Its output in discrete time domain is given by,

$$y(k,l) = \frac{1}{N}\sum_{p=1}^{P} u(p,l)e^{(j2\pi(p-1)(k-1)/N)} \qquad k = 1...N \tag{3}$$

where input u dimension P is an integer power of two and the length N is set equal to P.

To generate a baseband OFDM symbol, a serial digitized data stream is first channel coded and then modulated using DQPSK modulation. These data symbols are converted from serial-to-parallel into N data constellation points before modulating the subcarriers using IFFT, where N is the number of IFFT points.

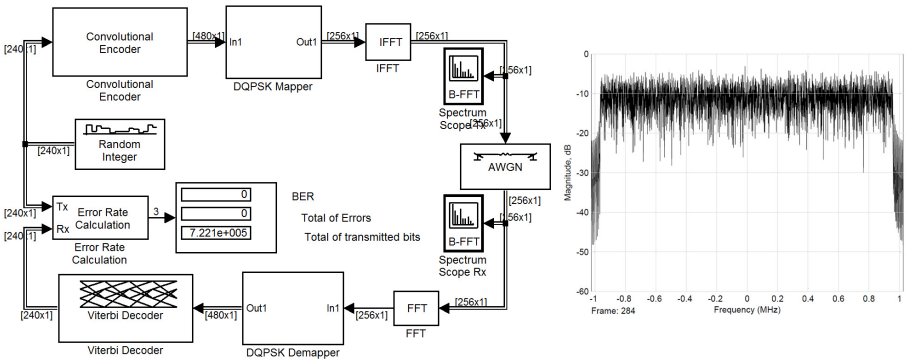


Fig. 3. Transmitter-receiver scheme using OFDM with DQPSK and signal spectrum

The receiver performs the exact opposite of the transmitter. The output of the FFT in frequency domain is given by,

$$y(k,l) = \sum_{p=1}^{P} u(p,l) e^{-(j2\pi(p-1)(k-1)/N)} \qquad k = 1...N \tag{4}$$

Here, transmission speed upon use of convolutional codification is double (3.84 Mbps) due to the increase of the number of bits per block to transmit (480 bits). Each carrier is modulated with DQPSK and we used a spectral efficiency of 2 bits/symbol, an IFFT of 256 points and 120 carriers (240 complex symbols equally distributed between the low and high parts of the spectrum). A total transmission speed of 1,920,000 symbols/s is reached. Transmission speed and flow for each carrier are 16,000 symbols/s and 32 kbps, respectively. Symbol period is 125 μs and separation between carriers (spectral resolution) is 8 kHz, resulting in a bandwidth (BW) of 0.96 MHz; see Figure 3. As is appreciated in Figure 3, a block that simulates a noisy channel was added to the modulation scheme so as to determine the minimum signal-to-noise relation (SNR) with a BER (Bit Error Rate) equal to zero, yielding a minimum SNR of 5 dB.

2.2 OFDM with QAM

This one is similar to the previous method in that it uses convolutional codification. The number of bits per block is double for a speed of 3.84 Mbps.

Each carrier is modulated using QAM and a spectral efficiency of 4 bits/symbol (with an IFFT of 128 points and 60 carriers). Total transmission speed is 960,000 symbols/s. Transmission speed per carrier is 16,000 symbols/s, for a flow of 64 kbps per carrier. Symbol period is 125 μs; 8 kHz is the separation between carriers for a bandwidth of 0.48 MHz. Under this modulation, BER equal to zero and a SNR minimum of 5 dB were achieved.

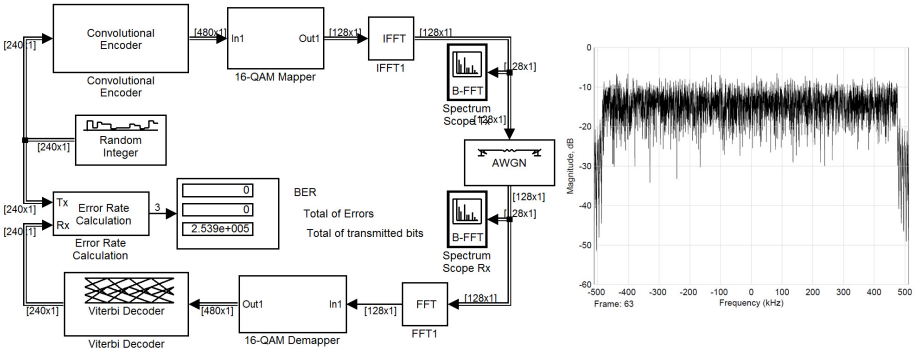


Fig. 4. Transmitter and receiver scheme for OFDM with 16-QAM and signal spectrum

The OFDM transmitter/receiver scheme using 16-QAM as a modulation technique can be observed in Figure 4 and the signal spectrum of the modulator output is also shown in this figure.

3 Results

Taking into account the results of the simulations, in order to increase OFDM spectral efficiency, it is important to use modulation schemes of a higher bits-symbol ratio, as can be appreciated in Figures 3 and 4. Because of this, when carriers modulate over 16-QAM, bandwidth decreases to 0.48 MHz, half of that achieved with DQPSK to obtain the OFDM signal (0.96 MHz).

According to the results of simulations, the scheme used with 16-QAM presents greater spectral efficiency than the one with DQPSK, with equal SNR value in both schemes without producing information loss, which is why the OFDM modulation scheme is more appropriate than 16-QAM for the design. Besides, since transmitters and receivers in antenna of cochlear implants are directional and separated by just a few millimeters, SNR in them should be a great deal higher than the minimum SNR resulting from the simulation, so as to obtain a reception without error.

4 Conclusions and Future Work

The aim of our research was fulfilled through the design and simulation of a telemetry system for a cochlear implant with high spectral efficiency, using OFDM. The design and simulation of the control system were achieved over FPGA. We managed to design and simulate an OFDM modulator supporting transmission speeds in the order of 3.84 Mbps within a bandwidth of 0.48 MHz, using 16-QAM and 0.96 MHz with DQPSK. It was verified that there is no difference regarding immunity to noise between OFDM using the DQPSK modulation

technique and OFDM using the QAM modulation technique. For all these reasons, the OFDM modulation scheme using QAM as a modulation technique was chosen for our design, since it is more spectral-efficient than DQPSK. With the use of convolutional codification as a means for error detection and correction, BER was kept equal to zero for signal relation noise values of up to 5 dB.

Acknowledgements. We would like to thank engineer Alexis Surez Mendoza for his contribution to the fulfillment of the present work.

References

1. Mai, S., Zhang, C., Dong, M., Wang, Z.: A cochlear system with implant DSP. In: IEEE International Conference on Acoustis, Speech and Signal Processing V, pp. 125–128 (2006)
2. Wilson, B.S.: Engineering Design of Cochlear Implants. In: Zeng, F.G., Popper, A.N., Fay, R.R. (eds.) Cochlear Implants. Auditory Prostheses and Electric Hearing. Springer (2004)
3. Bhoir, D.V., Panse, M.: Advances in Cochlear Implant Implementation. International Journal of Recent Trends in Engineering 2(8), 57–59 (2009)
4. Naghmouchi, F., Ghorbel, M., Hamida, A.B., Samet, M.: CMOS ASK System Modulation Dedicated to Cochlear Prothesis. In: IEEE First International Sysmposium on Control, Communications and Signal Processing, pp. 267–270 (2004)
5. Zeng, F.G., Rebscher, S., Harrison, W., Sun, X., Feng, H.: Cochlear Implants: System Design, Integration, and Evaluation. IEEE Reviews in Biomedical Engineering 1, 115–142 (2008)
6. Sawan, M., Hu, Y., Coulombe, J.: Wireless Smart Implants Dedicated to Multichannel Monitoring and Microstimulation. IEEE Circuits & Systems Magazine 5, 21–39 (2005)
7. Tang, Z., Smith, B., Schild, J.H., Peckham, P.H.: Data transmission from an implantable biotelemeter by load-shift keying using circuit configuration modulator. IEEE Transactions on Biomedical Engineering 42(5), 524–528 (1995)
8. Ghovanloo, M., Najafi, K.: A Wideband Frequency-Shift Keying Wireless Link for Inductively Powered Biomedical Implants. IEEE Trans. Circuits and Systems I 51(12), 2374–2383 (2004)
9. Martínez Rams, E.A., Cano Ortiz, S.D., Garcerán Hernández, V.: Cochlear Stimulator: Evaluation and Emulation Platform (2008)
10. Wise, K.D., Anderson, D.J., Hetke, J.F., Kipke, D.R., Najafi, K.: Wireless Implantable Microsystems: High-Density Electronic Interfaces to the Nervous System. In: Proceedings of the IEEE, vol. 92, pp. 76–97 (2004)
11. Yu, H., Najafi, K.: Low-power interface circuits for bio-implantable microsystems, Technical Digest. In: IEEE Int. Conf. Solid-State Circuits, San Francisco CA (2003)
12. Ghovanloo, M., Najafi, K.: A high-rate frequency shift keying demodulator chip for wireless biomedical implants. In: Proceedings of ISCAS 2003, vol. 5, pp. 45–48 (2003)
13. Hu, Y., Sawan, M.: A fully integrated low-power BPSK demodulator for implantable medical devices. IEEE Trans. Circ. Syst. 52, 2552–2562 (2005)
14. Hannan, M.A., Abbas, S.M., Samad, S.A., Hussain, A.: Modulation Techniques for Biomedical Implanted Devices and Their Challenges. Sensors 12, 297–319 (2012)

15. Luo, Z., Sonkusale, S.: A Novel BPSK Demodulator for Biological Implants. IEEE Trans. Circuits and Systems-I: Regular Papers 55(6), 1478–1484 (2008)
16. Deng, S., Hu, Y., Sawan, M.: A High Data Rate QPSK Demodulator for Inductively Powered Electronics Implants. In: IEEE International Symposium on Circuits and Systems, pp. 2577–2580 (2006)
17. Kiourti, A., Demosthenous, A.: BER Performance of a BPSK Biomedical Telemetry System under Varying Coupling and Loading Conditions. In: Lin, J. (ed.) MobiHealth 2010. LNICST, vol. 55, pp. 144–150. Springer, Heidelberg (2011)
18. Asgarian, F., Sodagar, A.M.: A Low-Power Noncoherent BPSK Demodulator and Clock Recovery Circuit for High-Data-Rate Biomedical Applications. In: 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4840–4843 (2009)
19. Sonkusale, S., Luo, Z.: A Complete Data and Power Telemetry System Utilizing BPSK and LSK Signaling for Biomedical Implants. In: 30th Annual International IEEE EMBS Conference, pp. 3216–3219 (2008)
20. Ackermann, D.M.: High Speed Transcutaneous Optical Telemetry Link. Case Western Reserve University (2007)
21. Neihart, N.M., Harrison, R.R.: A low-power FM transmitter for use in neural recording applications. In: Proc. IEEE Engineering in Medicine and Biology Conf., pp. 2117–2120 (2004)
22. Proakis, J.G.: Digital Communications, 4th edn. McGraw-Hill, Inc., New York (2000)
23. Peterson, L.L., Davie, B.S.: Computer Networks: A Systems Approach, 5th edn. The Morgan Kaufmann Series in Networking (2011)
24. Cimini, L.J.: Analysis and Simulation of a Digital Mobile Channel using Orthogonal Frecuency Divison Multiplexing. IEEE Transactions on Communications 33(7), 665–675 (1985)
25. Speth, M., Fechtel, S.A., Fock, G., Meyr, H.: Optimum Receiver Design for Wireless Broad-Band Systems Using OFDM. IEEE Transactions on Communications 47(11) (1999)

Deficitary Nervous Excitability and Subjective Contraction of Time: Time-Dispersive Model

Isabel Gonzalo-Fonrodona[1] and Miguel A. Porras[2]

[1] Departamento de Óptica, Facultad de Ciencias Físicas,
Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040-Madrid, Spain
igonzalo@fis.ucm.es
[2] Departamento de Física Aplicada y Grupo de Sistemas Complejos,
Universidad Politécnica de Madrid, Rios Rosas 21, 28003-Madrid, Spain
miguelangel.porras@upm.es

Abstract. A time-dispersive model of the nervous tissue is used to explain temporal aspects of perception, in relation to the observations and research made by J. Gonzalo [*Dinámica Cerebral*. Red Temát. Tec. Comput. Natural/Artificial, Univ. Santiago de Compostela, Spain 2010] on human cases with deficitary nervous excitation due to loss of neural mass. In these cases, the cerebral system is less excitable, has a lower reaction time and a slower decay of the cerebral excitation than in a normal case. The model considers the macroscopic excitation response of the neural network involved in the cerebral integration, and contains essential ingredients such as permeability to the excitation of the network and its reaction time. This model accounts for the observed shortening of the perceived duration of a stimulus, the perceived reduction of the lapse between two events, and the lower discrimination between repeated stimuli, when the network of the cerebral system is deficitary. Temporal summation is involved in these effects. The deficit of excitation makes the system more permeable to perception improvement by the presence of other type of stimulus (cross-modal effect due to multisensory integration), or by increasing the intensity of the stimulus, i. e., the cerebral excitation, in which case the subjective contraction of time tends to disappear.

Keywords: time perception, temporal discrimination, nervous excitation, temporal summation, chronaxia, reaction time.

1 Introduction

Since very early it was pointed out that a kinetic stimulus is perceived quicker under low illumination [1], or if it is perceived in the periphery of the visual field [2]. This effect could be interpreted as a shortening of the perceived time for the mobil to perform a given trajectory [1]. Also, for targets of constant size, temporal resolution between repeated stimuli decreases with eccentricity in the visual field [3–6]. These examples would illustrate a lowering of the nervous excitation that results in a decrease in time discrimination or "temporal acuity".

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 368–375, 2013.

This type of phenomenology is extremely pronounced in human cases characterized by Gonzalo [1, 7] as *central syndrome* cases, suffering from a marked loss of cerebral excitation due to a unilateral parieto-occipital cortex lesion in a rather unspecific (or multisensory) zone equidistant from the visual, tactile and auditory primary areas. They show bilateral and symmetric multisensory disorders.

A remarkable point is that this syndrome is similar to a normal case but in a smaller scale with respect to the cerebral excitability. This scale reduction leads to a desynchronization between different sensory qualities due to the loss of the different functions in a well-defined order according to their different excitability demands [1, 7, 8]. This research has been analyzed by different authors [9–20].

Spatial and temporal acuity, as well as movement perception requiring high spatial and temporal discrimination, are complex functions that would need high nervous excitation degree (and cerebral integration), and therefore they are among the first functions lost when the excitability of the neural network decreases [1, 8].

In cases with central syndrome and under low intensity of the stimulus, the temporal interval between successive stimuli is perceived as shortened, the flicker fusion frequency is lower than in normal, and a movement is perceived quicker than it actually is. This occurs in visual, tactile and auditory systems and is accompanied by many other disorders such as tilt or inversion perception of the stimulus, diminution of its size and loss of its shape and color [1, 7, 17, 19].

As the deficit of excitation is greater, the cerebral system becomes more permeable to temporal summation and also to cross-modal effects by multisensory integration. In a cross-modal effect, the perception of a stimulus improves markedly under the presence of another type of stimulus that supplies the excitation deficit, the system becoming more excitable and rapid [1, 7]. A functional gradients model of the cortex accounts for multisensory interactions as an inherent component of functional brain organization [7, 8]. This is in close relation and in agreement with other works [21–28].

Here we try to explain by means of a simple model the seemingly paradoxical fact that a deficitary cerebral system with low excitability and slow response perceives events as if they were faster than they actually are. For this, we focus on macroscopic properties of the nervous tissue such as the permeability to the excitation and the time response. These properties account for the capability of temporal summation, which is magnified in slow systems: In them, there is a long delay between the stimulus and the initiation of perception (long latency), and also a slow decay of the sensation (longer persistence) and, if a second stimulus arrives before the excitation of the first one has completely fallen down, there is an effect of summation in the cerebral excitation. The same occurs for a train of intermittent stimuli, so that it is possible to reach the excitation threshold to produce a sensorial perception despite a single one could be unable to do it.

In a previous work [16] we proposed a time-dispersive model involving nonlocal temporal effects to account for experimental intensity-duration data of electrical stimuli to produce minimum sensation. Here we improve that model to account

for the shortening of the duration of a stimulus, the shortening of the time interval between two stimuli, and the temporal discrimination between consecutive stimuli. The model establishes a relation between these temporal aspects and the macroscopic excitation response of the neural system, and can be applied to stimuli of different nature, as visual, tactile or auditive. Although a space-dispersive behaviour can also be considered, we restrict ourselves to temporal aspects only.

2 Model and Results

We consider a time-dispersive model where the cerebral excitation response at a time t depends on the stimulus at all previous times. If a stimulus $S(t)$ comes from the outside and acts on the cerebral system, the cerebral excitation $E(t)$ at a time t is the superposition of the effects produced at that instant and all previous instants. In a linear approximation, this can be expressed by

$$E(t) = \int_{-\infty}^{+\infty} \chi(\tau) S(t-\tau) d\tau \,, \tag{1}$$

where τ is the delay with respect to an instant t of the stimulus. Causality imposes $\chi(\tau) = 0$ for $\tau < 0$. The magnitude $\chi(\tau)$ is related to the capability of the system to become excited, and could also be interpreted as the excitation under a delta pulse stimulus at $t = 0$, i. e.,

$$E_\delta(t) = \int_{-\infty}^{+\infty} \chi(\tau) \delta(t-\tau) d\tau = \chi(t). \tag{2}$$

A simple model for $\chi(\tau)$ that accounts for the most relevant aspects of the excitation response to the stimulus is

$$\chi(\tau) = \chi_0 \, a\tau \, e^{-a\tau} \theta(\tau) \,, \tag{3}$$

where $\theta(\tau)$ is the Heaviside step function. The excitation $E(t)$ represents a macroscopic measure of the cerebral network excitation due to a stimulus, and involves the propagation of many induced nervous impulses. In Eq. (3), the constant χ_0 measures the excitation permeability, and the factor $a\tau$ describes approximately the initial growth of the response. This growth is followed by an exponential decay of the excitation $e^{-a\tau}$. The parameter a is directly related to the velocity of the system, i. e., with the chronaxia of the tissue (related to $1/a$). If the chronaxia is high (small a), the system is slow and the excitation increases and decays slowly. The parameters χ_0 and a characterize in this simple model the neural network for a given sensory function.

In what follows we compare the nervous excitation response of two cases: a normal human (N), and a deficitary one (M), the latter being a typical case of central syndrome case [1], characterized by lower values of χ_0 and a than in the normal case. The relative values of χ_0 and a used here for N and M have been

extracted from a previous work [16], where a time-dispersive model was used to fit the observed threshold intensity-duration data of electrical stimulation of the retina to obtain minimum sensation [1]. According to these data $\chi_{0,M}/\chi_{0,N} = 0.15$ and $a_M/a_N = 0.25$. A reasonable assumption is that the *cerebral* excitation threshold E_{th} to obtain minimum sensation is the same for a normal case (N) and for the deficitary case (M).

In Fig. 1(a), the response to a delta pulse stimulus at $t = 0$ is shown for cases N and M. For a particular threshold value E_{th} to perceive the sensation, it could happen that the threshold is not reached, as in central syndrome cases that do not perceive an isolated very short stimulus [1].

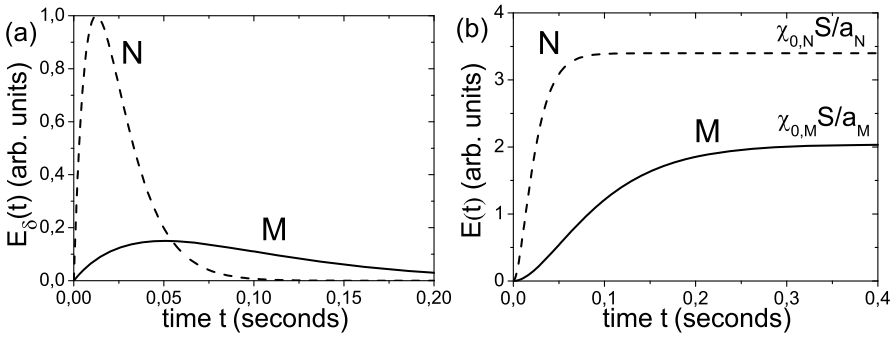


Fig. 1. Nervous excitation response $E(t)$ in a normal case N (dashed line) and a deficitary case M (solid line) to (a) a delta pulse stimulus and (b) a constant stimulus $S = 100$ (arb. units). For N, $a_N = 80$ s^{-1}, $\chi_{0,N} = 2.72$ (arb. units); for M, $a_M = 20$ s^{-1}, $\chi_{0,M} = 0.41$.

For a constant stimulus, starting from $t = 0$, i. e., $S(t) = S\theta(t)$, Eq. (1) with Eq. (3) yields

$$E(t) = \frac{\chi_0 S}{a}[1 - (1 + at)e^{-at}]\theta(t)\,, \tag{4}$$

where the excitation saturates at the value $(\chi_0 S/a)$ for large times, as shown for cases N and M in Fig. 1(b).

On this basis, we analyze the cerebral excitation for different stimuli of interests. In all them, the intensity of the stimulus S and the threshold value E_{th} for perception are chosen so that the response of N is well above E_{th}, while the response of the deficitary and slow case M is above but close to E_{th}.

We first consider a single constant stimulus of magnitude S acting during a time interval $\Delta t = 0.3$ s, i. e., $S(t) = S[\theta(t) - \theta(t - \Delta t)]$. From Eqs. (1) and (3), the neural excitation response is

$$E(t) = \frac{\chi_0 S}{a}[1-(1+at)e^{-at}]\theta(t) - \frac{\chi_0 S}{a}[1-(1+a(t-\Delta t)e^{-a(t-\Delta t)}]\theta(t-\Delta t), \tag{5}$$

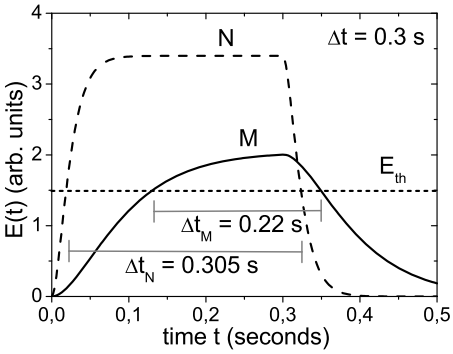


Fig. 2. Nervous excitation response to a constant stimulus starting from $t = 0$, magnitude $S = 100$, and duration $\Delta t = 0.3$ s, for N (dashed line) and for M (solid line)

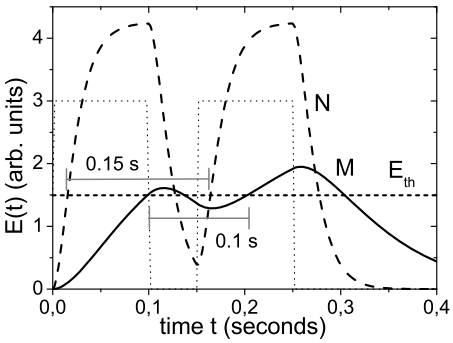


Fig. 3. Nervous excitation response $E(t)$ to two consecutive constant stimuli (thin dotted line) for N case (dashed line) and for case M (solid line)

which is represented in Fig. 2 for N and M. The excitation threshold E_{th} to perceive sensation is indicated by a dotted horizontal line. As seen, if the excitation reached is much greater than the excitation threshold E_{th}, the perceived duration of the stimulus is very close to the actual duration of the stimulus (the perceived duration is $\Delta t_N = 0.305$ s for case N in the example). In contrast, if the excitation reached is only slightly above the threshold for perception, as in case M, the perceived duration $\Delta t_N = 0.22$ s is significantly shorter than the actual duration of the stimulus.

We now consider the nervous excitation to two constant stimuli. The time interval between the initiation of the two stimuli is 0.15 s, as illustrated by the thin dotted line in Fig. 3. The excitation $E(t)$ can be similarly evaluated from Eq. (1) and is represented in Fig. 3 for cases N (dashed curve) and M (solid curve). A summation effect that impedes a good resolution between the two pulses can be appreciated. For N, excitation is high enough above the threshold

so that the perceived interval between the initiation of the two stimuli is quite similar to its actual value, but for M, this time interval is significantly shortened. In other words, when the excitation is close to the threshold for perception, the lapse of time between two events is perceived as shorter.

Finally, we apply the model to a series of identical consecutive stimuli. In this case, multiple situations can appear depending on the relative values of the parameters. As a relevant example, we consider the case that the duration of each pulse stimulus is 0.04 seconds, and the time interval between the initiation of two consecutive stimuli is 0.06 seconds (thin dotted lines in Fig. 4). This case evidences the effects of temporal summation in the excitation for N and M (dashed and solid curves). For N, summation causes the excitation to be permanently above the threshold E_{th} for perception, i. e., the intervals without stimulation are not perceived. The result is a flickering with rather pronounced oscillations. For M, summation causes the excitation to exceed the threshold only after the third pulse, and the excitation becomes almost continuous without oscillations, i.e., the flicker fusion frequency is already attained for this case.

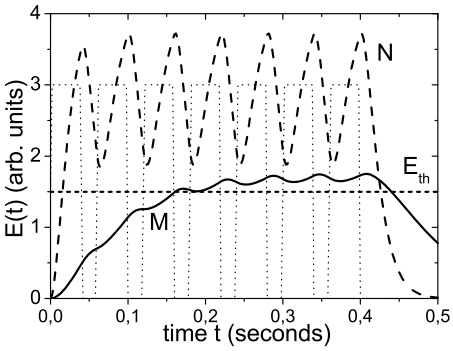


Fig. 4. Nervous excitation response to a train of identical intermittent stimulus (thin dotted line) for normal N case (dashed line) and deficitary M case (solid line)

3 Conclusion

Using a simple time-dispersive model, we have compared the different nervous excitation response to different type of temporal stimuli for a normal case and a deficitary case, the latter characterized by a lower excitability and slower reaction velocity than the normal one due to a lost of active neural mass. Stimuli can be visual, tactile or auditive. Data for these two cases have been taken from a previous works [1, 16]. The model considers the macroscopic excitation response of the neural network involved in the cerebral integration, contains basic ingredients such as permeability to excitation as well as its reaction time, and accounts for the main features of perception in relation to time discrimination or duration of events. In agreement with the observations [1, 7], the model describes the shortening of the perceived duration of a stimulus, the perceived

reduction of the time interval between two events, and the lower discrimination between successive stimuli, when the network of the cerebral system is deficitary in excitability and has slower temporal response than in a normal case.

This simple analysis admits obvious improvements as the introduction of non-linear effects that could explain the dependence of the subjective contraction of time with the intensity of the stimulus [1, 7, 29]. This is the subject of future work.

Acknowledgements. M. A. P. acknowledges financial support from Project MTM2012-39101-C02-01 of Ministerio de Economía y Competitividad of Spain.

References

1. Gonzalo, J.: Dinámica cerebral. La actividad cerebral en función de las condiciones dinámicas de la excitabilidad nerviosa. Consejo Superior de Investigaciones Científicas, Inst. S. Ramón y Cajal, Madrid, vol. I 1945, 1950. Enlarged edition: Gonzalo, J.: Dinámica cerebral. Red Temática en Tecnologías de Computación Natural/Artificial y Universidad de Santiago de Compostela, Spain (2010)
2. Exner, S.: Ein Versuch liber die Netzhaut Peripherie als Organ der Wahrnehmung von Bewegungen. Pflügers Archiv für die Gesamte Physiologie des Menschen und der Tiere 38, 215–218 (1886)
3. Poggel, D.A., Strasburger, H.: Visual perception in space and time. Mapping the visual field of temporal resolution. Acta Neurobiologiae Experimentalis 64, 427–436 (2004)
4. Poggel, D.A., Calmanti, C., Treutwein, B., Strasburger, H.: The Tölz Temporal Topography Study: Mapping the visual field across the life span: Part I. The topography of light detection and temporalinformation processing. Attention, Perception and Psychophysicss 74, 1114–1132 (2012)
5. Poggel, D.A., Treutwein, B., Calmanti, C., Strasburger, H.: Increasing the temporal grain: Double-pulse resolution is affected by the size of the attention focus. Vision Research 46, 2998–3008 (2006)
6. Strasburger, H., Rentschler, I., Jüttner, M.: Peripheral vision and pattern recognition: review. Journal of Vision 11(5), 13, 1–82 (2011)
7. Gonzalo, J.: Las funciones cerebrales humanas según nuevos datos y bases fisiológicas: Una introducción a los estudios de Dinámica Cerebral. Trab. Inst. Cajal Invest. Biol. 44, 95–157 (1952); Included as Gonzalo, J.: Dinámica Cerebral, suplem. I. Red Temática en Tecnologías de Computación Natural/Artificial y Universidad de Santiago de Compostela, Spain (2010)
8. Gonzalo, J.: Dinámica Cerebral. Red Temática en Tecnologías de Computación Natural/Artificial y Universidad de Santiago de Compostela, Spain (2010)
9. Bender, M.B., Teuber, H.L.: Neuro-ophthalmology. In: Spiegel III, E.A. (ed.) Progress in Neurology and Psychiatry, ch. 8, pp. 163–182 (1948)
10. Critchley, M.D.: The Parietal lobes, Arnold, London (1953)
11. Delgado, A.E.: Modelos Neurocibernéticos de Dinámica Cerebral. Ph.D.Thesis. E.T.S. de Ingenieros de Telecomunicación. Univ. Politécnica, Madrid (1978)
12. Mira, J., Delgado, A.E., Moreno-Díaz, R.: The fuzzy paradigm for knowledge representation in cerebral dynamics. Fuzzy Sets and Systems 23, 315–330 (1987)

13. Mira, J., Manjarrés, A., Ros, S., Delgado, A.E., Álvarez, J.R.: Cooperative Organization of Connectivity Patterns and Receptive Fields in the Visual Pathway: Application to Adaptive Thresholdig. In: Sandoval, F., Mira, J. (eds.) IWANN 1995. LNCS, vol. 930, pp. 15–23. Springer, Heidelberg (1995)
14. Gonzalo, I., Gonzalo, A.: Functional gradients in cerebral dynamics: The J. Gonzalo theories of the sensorial cortex. In: Moreno-Díaz, R., Mira, J. (eds.) Brain Processes, Theories and Models. An Int. Conf. in Honor of W.S. McCulloch 25 Years After his Death, pp. 78–87. MIT Press, Massachusetts (1996)
15. Gonzalo, I.: Allometry in the J. Gonzalo's model of the sensorial cortex. In: Cabestany, J., Mira, J., Moreno-Díaz, R. (eds.) IWANN 1997. LNCS, vol. 1240, pp. 169–177. Springer, Heidelberg (1997)
16. Gonzalo, I., Porras, M.A.: Time-dispersive effects in the J. Gonzalo's research on cerebral dynamics. In: Mira, J., Prieto, A.G. (eds.) IWANN 2001. LNCS, vol. 2084, pp. 150–157. Springer, Heidelberg (2001)
17. Gonzalo-Fonrodona, I.: Inverted or tilted perception disorder. Revista de Neurología 44, 157–165 (2007)
18. Gonzalo-Fonrodona, I., Porras, M.A.: Physiological Laws of Sensory Visual System in Relation to Scaling Power Laws in Biological Neural Networks. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4527, pp. 96–102. Springer, Heidelberg (2007)
19. Gonzalo-Fonrodona, I.: Functional gradients through the cortex, multisensory integration and scaling laws in brain dynamics. Neurocomputing 72, 831–838 (2009)
20. Gonzalo-Fonrodona, I., Porras, M.A.: Scaling effects in crossmodal improvement of visual perception by motor system stimulus. Neurocomputing (2012), doi:10.1016/j.neucom.2012.06.047
21. Pascual-Leone, A., Hamilton, R.: The metamodal organization of the brain. In: Casanova, C., Ptito, M. (eds.) Progress in Brain Research, vol. 134, pp. 1–19. Elsevier, Amsterdam (2001)
22. Pascual-Leone, A., Amedi, A., Fregni, F., Merabet, L.: The plastic human brain cortex. Ann. Rev. Neurosci. 28, 377–401 (2005)
23. Ghazanfar, A.A., Schroeder, C.E.: Is neocortex essentially multisensory? Trends. Cogn. Sci. 10, 278–285 (2006)
24. Laurienti, P.J., Burdette, J.H., Maldjian, J.A., Wallace, M.T.: Enhanced multisensory integration in older adults. Neurobiol. Aging 27, 1155–1163 (2006)
25. Martuzzi, R., Murray, M.M., Michel, C.M., Thiran, J.P., Maeder, P.P., Clarke, S., Meuli, R.A.: Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb. Cortex 17, 1672–1679 (2007)
26. Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault Jr., T.J., Rowland, B.A.: Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. Exp. Brain Res. 198, 113–126 (2009)
27. Hertz, U., Amedi, A.: Disentangling unisensory and multisensory components in audiovisual integration using a novel multifrequency fMRI spectral analysis. NeuroImage 52, 617–632 (2010)
28. Shams, L., Kim, R.: Crossmodal influences on visual perception. Phys. Life. Rev. 7, 269–284 (2010)
29. Lakhani, B., Vette, A.H., Mansfield, A., et al.: Electrophysiological Correlates of Changes in Reaction Time Based on Stimulus Intensity. PLoS ONE 7(5), e36407 (2012), doi:10.1371/journal.pone.0036407

Hybrid Tabu Search for Fuzzy Job Shop

Juan José Palacios[1], Jorge Puente[1], Inés González-Rodríguez[2], and Camino R. Vela[1]

[1] A.I. Centre and Department of Computer Science, University of Oviedo, Spain
{palaciosjuan,puente,crvela}@uniovi.es
http://di002.edv.uniovi.es/iscop
[2] Department of Mathematics, Statistics and Computing, University of Cantabria, Spain
ines.gonzalez@unican.es

Abstract. We consider the fuzzy job shop scheduling problem, which is a variant of the well-known job shop problem, with uncertainty in task durations that we model using fuzzy numbers. We propose a tabu search algorithm for minimising the expected makespan based on reversing arcs within critical blocks. We test the algorithm and then combine it with a genetic algorithm from the literature so we can observe the synergy effect, obtaining better results with the hybrid algorithm than with its components by separate. Finally we compare our hybrid algorithm with a memetic algorithm from the literature and show that even in similar times, our method is better in terms of expected makespan.

1 Introduction

Scheduling problems have formed an important body of research during the last decades as they are present in multiple applications in industry, finance and science [15]. Part of that research is focused on dealing with the uncertainty and vagueness pervading real-world situations [9]. Among the different ways of representing the uncertainty, fuzzy sets have emerged as a very interesting tool and have been extensively used in different manners, ranging from representing incomplete or vague states of information to using fuzzy priority rules with linguistic qualifiers or preference modelling [4],[18].

In deterministic scheduling the complexity of problems such as shop problems means that practical approaches to solving them usually involve heuristic strategies: simulated annealing, genetic algorithms, local search, etc. [2]. Some attempts have been made to extend these heuristic methods to the case where uncertain durations are modelled via fuzzy intervals, among others: a genetic algorithm is hybridised with a local search procedure in [10] for the flow shop problem, and in [13] a particle swarm is proposed to solve the open shop problem. For the job shop with different optimisation criteria, we find a neural approach [19], genetic algorithms [17],[14], simulated annealing [5], genetic algorithms hybridised with local search [6] or particle swarm optimisation [12].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 376–385, 2013.

In this paper, we intend to advance in the study of local search methods to solve the fuzzy job shop problem with expected makespan minimisation, denoted $J|fuzz\,p_i|E[C_{max}]$ according to the three field notation. We shall propose a new local search algorithm and see how it can be combined with a genetic algorithm to improve the quality of the best solutions found so far.

2 The Fuzzy Job Shop Scheduling Problem

The *job shop scheduling problem*, also denoted *JSP*, consists in scheduling a set of jobs $\{J_1, \ldots, J_n\}$ on a set of physical resources or machines $\{M_1, \ldots, M_m\}$, subject to a set of constraints. There are *precedence constraints*, so each job J_i, $i = 1, \ldots, n$, consists of m tasks $\{\theta_{i1}, \ldots, \theta_{im}\}$ to be sequentially scheduled. Also, there are *capacity constraints*, whereby each task θ_{ij} requires the uninterrupted and exclusive use of one of the machines for its whole processing time. A feasible schedule is an allocation of starting times for each task such that all constraints hold. The objective is to find a schedule which is *optimal* according to some criterion, most commonly that the *makespan* is minimal.

2.1 Uncertain Durations

In real-life applications, it is often the case that the exact time it takes to process a task is not known in advance, and only some uncertain knowledge is available. Such knowledge can be modelled using a *triangular fuzzy number* or TFN, given by an interval $[n^1, n^3]$ of possible values and a modal value n^2 in it. For a TFN N, denoted $N = (n^1, n^2, n^3)$, the membership function takes the following triangular shape:

$$\mu_N(x) = \begin{cases} \frac{x-n^1}{n^2-n^1} & : n^1 \leq x \leq n^2 \\ \frac{x-n^3}{n^2-n^3} & : n^2 < x \leq n^3 \\ 0 & : x < n^1 \text{ or } n^3 < x \end{cases} \tag{1}$$

In the job shop, we essentially need two operations on fuzzy numbers, the sum and the maximum. These are obtained by extending the corresponding operations on real numbers using the *Extension Principle*. However, computing the resulting expression is cumbersome, if not intractable. For the sake of simplicity and tractability of numerical calculations, we follow [5] and approximate the results of these operations, evaluating the operation only on the three defining points of each TFN. It turns out that for any pair of TFNs M and N, the approximated sum $M + N \approx (m^1 + n^1, m^2 + n^2, m^3 + n^3)$ coincides with the actual sum of TFNs; this may not be the case for the maximum $\max(M, N) \approx (\max(m^1, n^1), \max(m^2, n^2), \max(m^3, n^3))$, although they have identical support and modal value.

The membership function of a fuzzy number can be interpreted as a possibility distribution on the real numbers. This allows to define its expected value [11], given for a TFN N by $E[N] = \frac{1}{4}(n^1 + 2n^2 + n^3)$. It coincides with the neutral

scalar substitute of a fuzzy interval and the centre of gravity of its mean value [4]. It induces a total ordering $\leq_E$ in the set of fuzzy numbers [5], where for any two fuzzy numbers M, N $M \leq_E N$ if and only if $E[M] \leq E[N]$.

2.2 Fuzzy Job Shop Scheduling

A job shop problem instance may be represented by a directed graph $G = (V, A \cup D)$. V contains one node $x = m(i-1) + j$ per task θ_{ij}, $1 \leq i \leq n$, $1 \leq j \leq m$, plus two additional nodes 0 (or *start*) and $nm+1$ (or *end*), representing dummy tasks with null processing times. Arcs in A, called *conjunctive arcs*, represent precedence constraints (including arcs from node *start* to the first task of each job and arcs form the last task of each job to node *end*). Arcs in D, called *disjunctive arcs*, represent capacity constraints; $D = \cup_{j=1}^{m} D_j$, where D_j corresponds to machine M_j and includes two arcs (x, y) and (y, x) for each pair x, y of tasks requiring that machine. Each arc (x, y) is weighted with the processing time p_x of the task at the source node (a TFN in our case). A feasible task processing order σ is represented by a *solution graph*, an acyclic subgraph of G, $G(\sigma) = (V, A \cup R(\sigma))$, where $R(\sigma) = \cup_{j=1}^{m} R_j(\sigma)$, $R_j(\sigma)$ being a hamiltonian selection of D_j. Using forward propagation in $G(\sigma)$, it is possible to obtain the starting and completion times for all tasks and, therefore, the schedule and the makespan $C_{max}(\sigma)$.

The schedule will be fuzzy in the sense that the starting and completion times of all tasks and the makespan are TFNs, interpreted as possibility distributions on the values that the times may take. However, the task processing ordering σ that determines the schedule is crisp; there is no uncertainty regarding the order in which tasks are to be processed.

Given that the makespan is a TFN and neither the maximum nor its approximation define a total ordering in the set of TFNs, it is necessary to reformulate what is understood by "minimising the makespan". In a similar approach to stochastic scheduling, it is possible to use the concept of expected value for a fuzzy quantity and the total ordering it provides, so the *objective* is to minimise the expected makespan $E[C_{max}(\sigma)]$, a crisp objective function.

Another concept that needs some reformulation in the fuzzy case is that of criticality, an issue far from being trivial. In [5], an arc (x, y) in the solution graph is taken to be critical if and only if the completion time of x and the starting time of y coincide in any of their components. In [8], it is argued that this definition yields some counterintuitive examples and a more restrictive notion is proposed. From the solution graph $G(\sigma)$, three *parallel solution graphs* $G^i(\sigma)$, $i = 1, 2, 3$, are derived with identical structure to $G(\sigma)$, but where the cost of arc $(x, y) \in A \cup R(\sigma)$ in $G^i(\sigma)$ is p_x^i, the i-th component of p_x. Each parallel solution graph $G^i(\sigma)$ is a disjunctive graph with crisp arc weights, so in each of them a critical path is the longest path from node *start* to node *end*. For the fuzzy solution graph $G(\sigma)$, a path will be considered to be *critical* if and only if it is critical in some $G^i(\sigma)$. Nodes and arcs in a critical path are termed critical and a critical path is naturally decomposed into critical blocks, these being maximal subsequences of tasks requiring the same machine.

Generate an initial solution S
$S^* \leftarrow S$
$tabuList \leftarrow \emptyset$
while $\neg StoppingCriterion$ **do**
  $\Omega \leftarrow Neighbourhood(S)$
  $\Omega \leftarrow \Omega - \{\omega_i \in \Omega \mid \omega_i \in tabuList \wedge \neg aspiration(\omega_i)\}$
  $S = ChooseNeighbour(\Omega)$
  if $Cmax(S) <_E Cmax(S^*)$ **then**
    $S^* \leftarrow S$
  Update $tabuList$
return S^*;

Alg. 1. General schema for tabu search

In order to simplify expressions, we define the following notation for a feasible schedule. For a solution graph $G(\sigma)$ and a task x, let $P\nu_x$ and $S\nu_x$ denote the predecessor and successor nodes of x on the machine sequence (in $R(\sigma)$) and let PJ_x and SJ_x denote the predecessor and successor nodes of x on the job sequence (in A). The *head* of task x is the starting time of x, a TFN given by $r_x = \max\{r_{PJ_x} + p_{PJ_x}, r_{P\nu_x} + p_{P\nu_x}\}$, and the *tail* of task x is the time lag between the moment when x is finished until the completion time of all tasks, a TFN given by $q_x = \max\{q_{SJ_x} + p_{SJ_x}, q_{S\nu_x} + p_{S\nu_x}\}$.

3 Tabu Search for FJSP

Roughly speaking, a typical local search schema starts from a given solution, calculates its neighbourhood and then chooses a promising neighbour, which is usually the neighbour with the best value for the objective function. The chosen neighbour replaces the current solution and the process repeats until a stopping criterion is met. The algorithm finally returns the best solution found so far which in our case means that one with the smallest $E[C_{max}]$. In case we have two solutions A and B with the same $E[C_{max}]$ we use a ranking from [1] which chooses the solution with the minor modal value and, if the tie persists, then chooses the solution with the minimum support width.

The simplest local search schema is Hill Climbing, which moves from a solution to a neighbour only if the latter provides an improvement. This approach is fast but it gets easily stuck in local optima. To prevent this, tabu search allows a solution S to move to a non-improving neighbour. This usually makes the algorithm find better solutions at the cost of more evaluations, and in consequence, longer runtime. In our tabu search, starting from an initial solution S, the algorithm moves towards the neighbour with the best $E[C_{max}]$ value. However, in the case that all neighbours of S are worse than it, we may choose a neighbour S' and find at the next step that the best neighbour of S' is S again so we get trapped in a loop. To avoid this, tabu search uses a *tabu list* that stores forbidden solutions or forbidden movements, which are called *tabu*. In addition

to a tabu status, a so-called *aspiration criterion* is associated with each move, so if a movement satisfies the associated aspiration criterion, it is considered an admissible move even if it is in the tabu list. Algorithm 1 shows the general schema for the tabu search.

3.1 Neighbourhood

During the last years, many neighbourhoods have been used for solving the job shop problem with deterministic task durations. In particular, in [20], a neighbourhood structure is introduced based on reversing all the critical arcs in the disjunctive graph $G(\sigma)$. This is extended to the fuzzy framework in [5], where an arc (x, y) is taken to be critical in $G(\sigma)$ if exists $i = 1, 2, 3$ such that $r_x^i + p_x^i = q_y^i$. A second extension to the fuzzy case was proposed in [8], using the definition of criticality based on parallel solution graphs instead. As a consequence of the criticality definitions, the new neighbourhood is a proper subset of the previous one while still containing all the improving solutions. Additionally, all neighbours in this structure are feasible and the connectivity property holds: starting from any solution, it is possible to reach a given global optimum in a finite number of steps using this structure. In [6] a new neighbourhood is proposed, based on reversing only those critical arcs at the extreme of critical blocks of a single path. Although the connectivity property does not hold any more, it contains only feasible solutions and it proves to be a very efficient structure.

More recently, in [16] the authors propose a new neighbourhood for the fuzzy job shop inspired in the work from [3] for the deterministic problem based on also reversing adjacent arcs to (x, y) (machine predecessor $P\nu_x$ and successor $S\nu_y$) as explained in Definition 1.

Definition 1. *Let σ be a task processing order and let $v = (x, y)$ be an arc at the extreme of a critical block in the associated graph $G(\sigma)$. Then, the* neighbourhood structure $\mathcal{N}_3^R(\sigma)$ *is obtained as follows: if (x, y) is the only arc in the critical block, then (x, y) is reversed; if $P\nu_x$ is also critical (and $S\nu_y$ is not), then we consider all possible permutations of $(P\nu_x, x, y)$ where (x, y) is reversed; else, if $S\nu_y$ is critical, then we consider all possible permutations of $(x, y, S\nu_y)$ where (x, y) is reversed.*

In the proposed tabu search, only the best neighbour is of interest. Thus, a makespan lower bound may help find the best neighbour in less time discarding those which are far from being the best of the neighbourhood. Therefore we use the method proposed in [16] which calculates a lower bound for the fuzzy makespan of a $\mathcal{N}_3^R$ neighbour as the longest path that would pass throw the affected nodes if we performed the move. This method also allows to easily discard moves that yield to non-feasible solutions (notice that the neighbourhood itself could generate non-feasible solutions). Details on how neighbours are chosen using lower bounds are given in Algorithm 2.

Method $ChooseNeighbour(\Omega)$
$\omega^* \leftarrow emptySolution$
Compute the lower bound $estim(\omega_i)$ for each neighbour $\omega_i \in \Omega$
while $\Omega \neq \emptyset$ **do**
  $\omega_c \leftarrow \arg\min_{\omega_i \in \Omega}\{estim(\omega_i)\}$
  $\Omega \leftarrow \Omega - \{\omega_c\}$
  Evaluate ω_c updating heads and tails [6]
  if ω^* is empty or $C_{max}(\omega_c) <_E C_{max}(\omega^*)$ **then**
    $\omega^* \leftarrow \omega_c$
    $\Omega \leftarrow \Omega - \{\omega_i \mid estim(\omega_i) > E[C_{max}(\omega^*)]\}$
return ω^*;

Alg. 2. Neighbour choice

3.2 The Tabu List

In its conception, a tabu list is a set of forbidden solutions so the local search does not explore them any more. However, storing complete solutions and testing if a neighbour belongs to the list is too inefficient in terms of computational time. Usually, a tabu list stores the opposite of any move applied during the search to transform a solution into a new one, e.g. the case that we reverse the arc (x, y), we forbid the reversal of arc (y, x) by including it in the tabu list. This simple case cannot be trivially extended to the neighbourhood used herein, where a critical arc (x, y) can generate up to 5 different neighbours.

If we move to a neighbour generated from the critical arc (x, y), we propose to store the relative order before the change of the involved tasks and forbid any movement that yields to a solution with that relative order. For example, if we are in the state $(P\nu_x, x, y)$ and decide to move to $(y, P\nu_x, x)$, we store the first tuple and forbid any movement that generates any solution in which these 3 tasks are sorted as $(P\nu_x, x, y)$. We also introduce a parameter $tabuSize$ that determines the maximum size of the list. The behaviour of the list is FIFO, that is, if we need to introduce a new forbidden order in the list and the $tabuSize$ has been reached, we remove the oldest one.

A tabu move is allowed provided that it fulfills an aspiration criterion, in our case, that its expected makespan improves the best solution found so far, provided that there is no non-tabu neighbour with same or better quality.

3.3 Stopping Criterion

Unlike other methods like hill climbing, tabu search algorithms do not have a well-established stopping criterion. In the algorithm we propose, the tabu search runs for a number of iterations and finishes if the most recent improvement of the best solution has not occurred in the last $maxIter$ iterations.

4 Experimental Results

The purpose of this section is to provide an experimental evaluation of the proposed algorithm in combination with a genetic algorithm. To this end, as in [6] we shall use a set of instances generated by fuzzifying 12 benchmark problems for job shop: the well-known FT10 (size 10×10) and FT20 (20×5), and La21, La24, La25 (15×10), La27, La29 (20×10), La38, La40 (15×15), and ABZ7, ABZ8, ABZ9 (20×15), a set of 10 problems considered to be hard to solve for classical job shop. We use a fuzzy instance of each benchmark, generated following [5], so task durations become symmetric TFNs where the modal value is the original duration, thus ensuring that the optimal solution to the crisp problem provides a lower bound (LB) for the fuzzified version. This allows to measure the quality of solutions as a relative error (RE) with respect to that lower bound.

$$RE = \frac{\mid E[C_{max}] - LB \mid}{LB} \tag{2}$$

All the experiments reported in this section, correspond to a C++ implementation running on a PC with Xeon processor at 2,2Ghz and 24 Gb RAM running Linux (SL 6.0.1).

Local search algorithms are very sensitive to the starting solution. It is a common practice to run the algorithm several times using different starting solutions (multi-start LS) that could be generated randomly or using a heuristic to obtain good starting points. Here, the multi-start feature is achieved by combining the tabu search (TS) with a genetic algorithm (GA); the resulting method is denoted *HTS*. This kind of hybridisation generally improves the quality of those methods run independently as the GA keeps the quality of the solutions and the diversity, while the TS provides a deeper exploitation of the solutions. We apply the TS to every individual in the population right after its evaluation, resulting in a so-called *memetic algorithm*. In the GA, individuals are permutations with repetitions which are evaluated using a fuzzyfied G&T algorithm. To obtain the best possible performance, a parametric analysis (not reported here due to lack of space) was conducted. The resulting parameter values were: Population=100, JOX crossover with probability 0.9, selection with random pairs (all the individuals are paired), 4:2 tournament between parents and their offspring for the replacement, tabu list size = 8 and $maxIter = 10$ as stopping criterion for the search.

To estimate the number of generations needed by HTS to converge we run it 10 times on the fuzzy benchmark and record average objective values for the best individual in the population in each run and the average population quality. We conclude the algorithm needs 100 generations to converge. The parameter $maxIter$ for the TS has a special relevance for the algorithm behaviour as it defines the length of the TS and as a consequence the level of exploration around the initial solution. Figure 1 illustrates the algorithm's behaviour with varying values of $maxIter$. We can appreciate that the algorithm is scalable as it can reduce the relative error when given more time. Of course the longer the runtime is, the lower the error reduction is. However, choosing a value for the parameter is not easy, since it depends strongly on the available runtime. Being aware of the

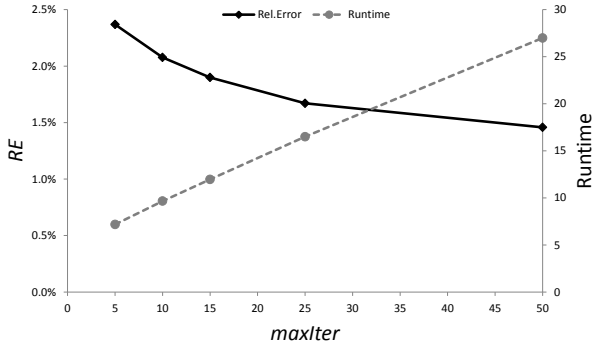


Fig. 1. Average *RE* depending on parameter *maxIter*

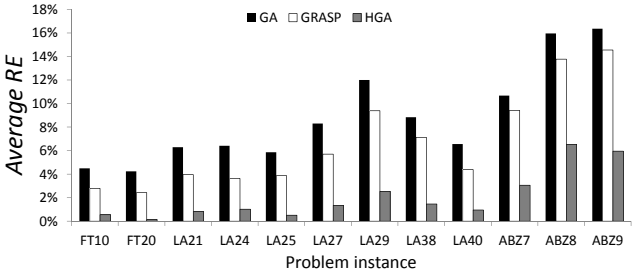


Fig. 2. Average RE obtained by GA, GRASP and HTS on every benchmark instances

sensitivity of this parameter, for this work we set it to 10 so we can compare HTS with other another method from the literature [16] in equal runtime conditions.

We run HTS on the benchmark 30 times and store the best and average solution values as well as average runtime across the 30 runs. In order to asses whether the combination of the TS with a GA adds any value to the TS alone, we have implemented a GRASP algorithm which generates random solutions using the same evaluation function as the GA and then applies the TS to them. For the sake of comparison, it generates 10.000 different starting solutions (HTS evaluates 100 populations of size 100). In addition, to assess the synergy effect we also run the GA with no local search, letting it evolve with the same population size (100) for the same time HTS takes to finish. Figure 2 shows there is a clear synergy effect; in terms of *RE*, the improvement of HTS with respect to the GA is 81,0% in average, being the minimum improvement 59,1% for ABZ8 and the maximum 96,3% for FT20. Despite the GRASP being better in *RE* than the GA, HTS outperforms it in 74,8%, with the improvement ranging from 52,5% (ABZ8) to 93,5% (FT20). In addition, it is remarkable that the runtime for the GRASP is more than 3 times (370%) longer than the runtime for the GA or HTS. This demonstrates the great relevance of the starting solutions for the TS.

Finally, we compare our algorithm with MA from the literature [16] denoted $MA-3$ therein, which combines a GA analogous to the one used here with

Table 1. Comparison with an MA from the literature

Method	Problem	*RE*		Time	Problem	*RE*		Time
		Best	Avg	(sec.)		Best	Avg	(sec.)
MA-3	FT10	0.30	1.06	2.8	LA29	1.45	3.53	8.4
HTS	(930)	0.30	**0.56**	4.2	(1130)	1.61	**2.54**	9.5
MA-3	FT20	0.04	0.84	3.8	LA38	0.61	2.53	9.0
HTS	(1165)	0.04	**0.16**	4.6	(1196)	0.61	**1.47**	10.1
MA-3	LA21	0.50	1.17	5.3	LA40	0.63	1.47	9.5
HTS	(1046)	0.33	**0.82**	6.0	(1222)	0.43	**0.96**	9.9
MA-3	LA24	0.61	1.21	5.0	ABZ7	2.21	4.00	15.9
HTS	(935)	0.56	**1.02**	5.9	(656)	2.06	**3.06**	16.7
MA-3	LA25	0.15	0.80	5.0	ABZ8	5.85	7.73	16.9
HTS	(977)	0.15	**0.51**	6.0	(645)	5.27	**6.53**	16.3
MA-3	LA27	0.30	2.01	9.0	ABZ9	4.88	7.36	17.2
HTS	(1236)	0.36	**1.35**	9.4	(661)	4.69	**5.95**	17.2

hill-climbing. Table 1 shows the best and average *RE* values obtained with both algorithms. We see that HTS is better than MA-3 in average for all the instances, having an average improvement of 33,9%. The highest improvement is in problem FT20 (81,2%) and the smallest is in problem LA24 (15,5%). If we perform an analysis per family, the best results are obtained on the FT problems, and the lesser improvement is for ABZ problems. We can also appreciate that the new algorithm takes less than an additional second in average in comparison with MA-3 in terms of computational time.

5 Conclusions

We have tackled a variant of the job shop scheduling problem where uncertainty in durations is modelled using triangular fuzzy numbers and where the objective is to minimise the expected makespan. We have proposed a TS algorithm and combined it with a GA in order to have better initial solutions and keep diversity. The experimental results have shown that this combination outperforms both the behaviour of the TS and the GA when they are run independently. Finally we have compared the new memetic algorithm HTS with a MA from the literature and obtained better results for all tested instances using similar computational times. Based in the promising results, in the future we intend to improve on the TS, starting by adapting the most competitive search algorithms for the problem with deterministic durations to the problem with uncertainty. We also intend to create additional harder benchmarks for the fuzzy job shop which provide greater room for improvement for future metaheuristics.

Acknowledgements. This research has been supported by the Spanish Government under research grants FEDER TIN2010-20976-C02-02 and MTM2010-16051.

References

1. Bortolan, G., Degani, R.: A review of some methods for ranking fuzzy subsets. In: Dubois, D., Prade, H., Yager, R. (eds.) Readings in Fuzzy Sets for Intelligence Systems, pp. 149–158. Morgan Kaufmann, Amsterdam (1993)
2. Brucker, P., Knust, S.: Complex Scheduling. Springer (2006)
3. Dell' Amico, M., Trubian, M.: Applying tabu search to the job-shop scheduling problem. Annals of Operational Research 41, 231–252 (1993)
4. Dubois, D., Fargier, H., Fortemps, P.: Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge. European Journal of Operational Research 147, 231–252 (2003)
5. Fortemps, P.: Jobshop scheduling with imprecise durations: a fuzzy approach. IEEE Transactions of Fuzzy Systems 7, 557–569 (1997)
6. González Rodríguez, I., Vela, C.R., Hernández-Arauzo, A., Puente, J.: Improved local search for job shop scheduling with uncertain durations. In: Proceedings of ICAPS-2009, pp. 154–161. AAAI Press, Thesaloniki (2009)
7. González Rodríguez, I., Vela, C.R., Puente, J.: A memetic approach to fuzzy job shop based on expectation model. In: Proceedings of IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2007, pp. 692–697. IEEE, London (2007)
8. González Rodríguez, I., Vela, C.R., Puente, J., Varela, R.: A new local search for the job shop problem with uncertain durations. In: Proceedings of ICAPS-2008, pp. 124–131. AAAI Press, Sidney (2008)
9. Herroelen, W., Leus, R.: Project scheduling under uncertainty: Survey and research potentials. European Journal of Operational Research 165, 289–306 (2005)
10. Ishibuchi, H., Murata, T.: A multi-objective genetic local search algorithm and its application to flowshop scheduling. IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviews 67(3), 392–403 (1998)
11. Liu, B., Liu, Y.K.: Expected value of fuzzy variable and fuzzy expected value models. IEEE Transactions on Fuzzy Systems 10, 445–450 (2002)
12. Niu, Q., Jiao, B., Gu, X.: Particle swarm optimization combined with genetic operators for job shop scheduling problem with fuzzy processing time. Applied Mathematics and Computation 205, 148–158 (2008)
13. José Palacios, J., González-Rodríguez, I., Vela, C.R., Puente, J.: Particle swarm optimisation for open shop problems with fuzzy durations. In: Ferrández, J.M., Álvarez Sánchez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2011, Part I. LNCS, vol. 6686, pp. 362–371. Springer, Heidelberg (2011)
14. Petrovic, S., Fayad, S., Petrovic, D., Burke, E., Kendall, G.: Fuzzy job shop scheduling with lot-sizing. Annals of Operations Research 159, 275–292 (2008)
15. Pinedo, M.L.: Scheduling, 3rd edn. Theory, Algorithms, and Systems. Springer (2008)
16. Puente, J., Vela, C.R., González-Rodríguez, I.: Fast local search for fuzzy job shop scheduling. In: Proceedings of ECAI 2010, pp. 739–744. IOS Press (2010)
17. Sakawa, M., Kubota, R.: Fuzzy programming for multiobjective job shop scheduling with fuzzy processing time and fuzzy duedate through genetic algorithms. European Journal of Operational Research 120, 393–407 (2000)
18. Słowiński, R., Hapke, M. (eds.): Scheduling Under Fuzziness, Studies in Fuzziness and Soft Computing, vol. 37. Physica-Verlag (2000)
19. Tavakkoli-Moghaddam, R., Safei, N., Kah, M.: Accessing feasible space in a generalized job shop scheduling problem with the fuzzy processing times: a fuzzy-neural approach. Journal of the Operational Research Society 59, 431–442 (2008)
20. Van Laarhoven, P., Aarts, E., Lenstra, K.: Job shop scheduling by simulated annealing. Operations Research 40, 113–125 (1992)

Statistical Characteristics of Portal Images and Their Influence in Noise Reduction

Antonio González-López[1], María-Consuelo Bastida-Jumilla[2], Jorge Larrey-Ruiz[2], and Juan Morales-Sánchez[2]

[1] Hospital Universitario Virgen de la Arrixaca, ctra. Madrid-Cartagena s/n, 30120 El Palmar (Murcia), Spain
[2] Departamento de Tecnologías de la Información y las Comunicaciones, Universidad Politéctnica de Cartagena. Cartagena, Spain
antonio.gonzalez7@carm.es, {mc.bastida,jorge.larrey,juan.morales}@upct.es

Abstract. Portal imaging is used in radiotherapy to asses the correct positioning of the patient before applying the treatment. Given the high energy particles used in portal image formation, portal image is intrinsically bound by low contrast and poor spatial resolution. The relevance of portal imaging in radiotherapy treatments and its common use justify efforts to improve its inherent low quality.

The knowledge of the statistical properties of both image and noise is essential in order to develop suitable processing algorithms to clean the image. The aim of this paper is to show how the statistical characteristics of the portal images and noise images generated in one of the portal imaging systems most widely deployed, can be exploited to improve the quality of noisy portal images through efficient denoising methods.

An ensemble of portal images is used to investigate their statistical characteristics. In the case of noise, a process of averaging and subtraction of the mean is used to extract noise images.

The distribution found for the noise is clearly Gaussian, in both the spatial and the wavelet domain. The curves for the noise show a parabolic shape in the semi-log graphs across the different scales, which translates into Gaussian character in the transformed domain. On the other hand, the probability density functions (pdf's) for portal images show large tails.

Wavelet thresholding takes advantage of the different statistical features found for noise and signal. In the present work wavelet thresholding is compared to Wiener filtering, and the assesment of the denoised image is carried out by means of the peak signal to noise ratio PSNR and the structural similarity index SSMI.

Thresholding the wavelet coefficients of the noisy image gives better denoising results for both figures of merit (PSNR and SSIM) than the Wiener filter in all the analysed cases. Furthermore, the differences between the methods increase as the noise increases. *abstract* environment.

Keywords: portal image, image statistics, wavelet processing, denoising.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 386–395, 2013.

1 Introduction

Portal imaging plays a crucial role in radiotherapy. In external radiotherapy a simulation is carried out prior to treatment. During simulation the geometry of the patient is explored to determine a convenient treatment strategy. During treatment the geometric conditions used for the treatment simulation must be reproduced. One of the most widespread methods to guarantee this consistency is to compare images acquired under the treatment beam (portal images) with synthetical images generated during the simulation phase (digitally reconstructed radiography or DRR). By matching both images, the agreement to the geometric scenario in the treatment simulation (planning stage) is determined.

Portal images have a feature that makes them unique among other verification methods: the treatment beam itself forms the image, so that the agreement between the patient positioning system and the treatment beam is inherent.

However, the portal image quality is intrinsically bound by the low contrast and low spatial resolution inherent to high energy radiation sources (photons with energies within the range of MeV). The limiting factor in the portal images contrast is the dominant type of the radiation-matter interactions at the energy levels used in radiotherapy. X-ray attenuation is dominated by the Compton effect, and the probability of Compton interactions is highly dependent on the electron density of the material, unlike what happens with the photoelectric effect, which shows a strong dependence on atomic number. Given that anatomic structures generally show small variations in the electron density, the obtained contrast for the energy levels of therapy is much lower than for the energy levels of diagnostic [1]. The radiation used in radiotherapy has a great ability to penetrate matter. This fact reduces the probability of interaction with the detector, resulting in a poor detection efficiency [2]. On the other hand, the secondary particles created in the detector have a high energy, resulting in a certain range of movement of these particles within the detector, thus leading to an impoverishment of the spatial resolution. Spatial resolution is also affected by the relatively large size of the focal spot of the beam. The relevance of portal imaging in radiotherapy treatments and its common use justify efforts to improve its inherent low quality.

The knowledge of the statistical properties of both image and noise is essential in order to develop suitable processing algorithms to clean the image. Statistical properties may reveal key differences between images and noise in denoising methods, or may guide the design of a priori distributions in Bayesian methods. Since the beginning of the wavelet transform, it has been noticed that wavelet thresholding is of considerable interest in order to remove noise from signals and images. The outcome of this thresholding depends on the existing differences between the marginal distributions of the wavelet coefficients corresponding to signal and noise [3]. Alternative denoising methods which consider joint statistics of the subband coefficients have shown a better performance than those based on the marginal statistics of the wavelet coefficients [4,5].

There are many cases where the denoising methods take advantage of the underlying statistical characteristics of medical images. In the case of radiographic images, denoising methods using hidden Markov trees exploit the statistical properties of these images in the wavelet domain [6].

The aim of this paper is to show how the statistical characteristics of the portal images, and the noise generated in one of the portal imaging systems most widely deployed, can be exploited to improve the quality of noisy portal images through efficient denoising methods.

2 Materials and Methods

2.1 Materials

The portal imaging system used in this study is the Portal Vision aS500 from Varian Medical Systems. The aS500 is an amorphous Silicon flat panel detector with an array of 384x512 transistors. The pixel size is 0.784mm x 0.784mm and the system contains a metal plate and a phosphor screen to convert the X-ray photons into photons in the visible spectrum. The transistors of the array transform these photons into an electrical signal. The radiation beams used have a nominal energy of 6MV and are produced by a linear accelerator Clinac 2100 DHX (Varian Medical Systems).

2.2 Image and Noise Statistical Distributions

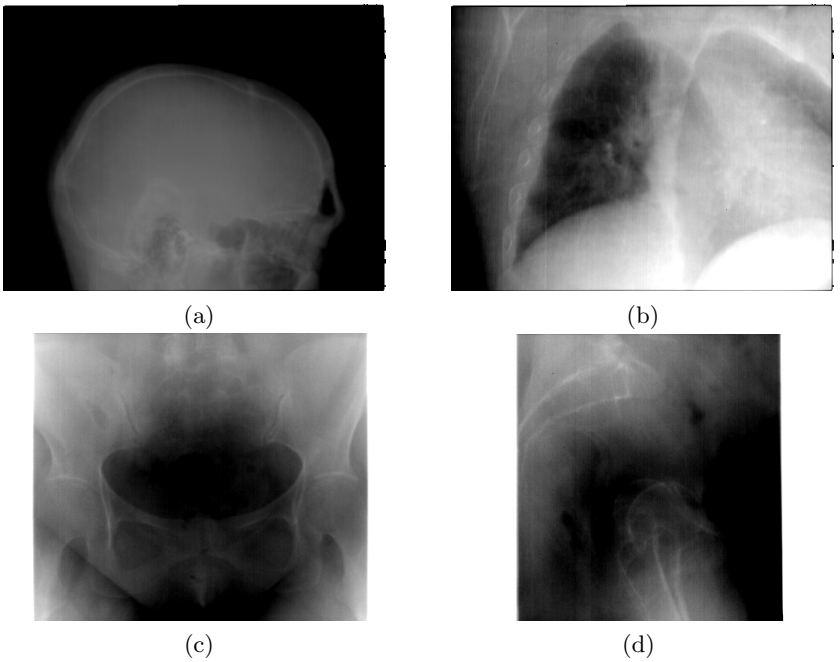
The present work studies the pobability density functions (pdf's) in the wavelet domain of both image and noise. The distribution of noise in the image domain is also revised. The pdf's are estimated by normalizing the computed histograms. With regard to the mathematical representation of noise, a classical additive model is considered, that is, an image is given by the addition of a signal plus noise.

The analysis of the statistical distributions of image and noise is motivated by the search of characteristics of the signal (anatomical information) in the image that do not appear in the noise. The description of these characteristics is the basis on which many image processing and noise reduction are based.

In order to determine the pdf's for the portal image, a total of 163 of these images, from different anatomical locations, were used. Figure 1 shows a sample of these images. A process of averaging and subtraction of the mean is used to achieve noise images. To study the power spectrum and the *pdf* of noise, a series of uniform images were employed (see Figure 2a). These images are obtained by irradiating a $30\,cm \times 30\,cm \times 15\,cm$ methacrylate phantom with a uniform beam. The average of 16 of these uniform images is subtracted from each of them to compute 16 noise images (Figure 2b).

2.3 Wavelet Analysis

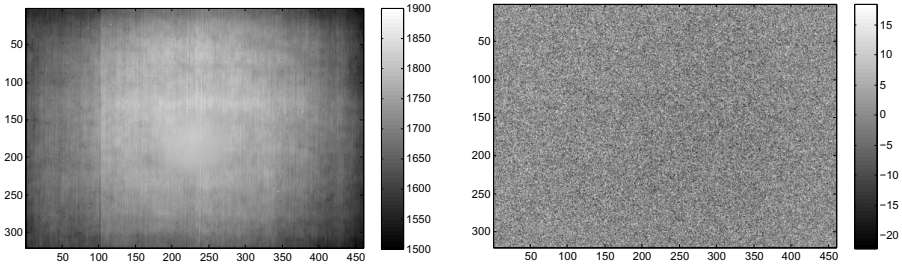
In order to estimate the pdf's of wavelet coefficients, the normalized histograms of such coefficients have to be previously computed. The decimated Haar transform to carry out the wavelet transformation. With this transformation we ensure



(a) (b)

(c) (d)

Fig. 1. A Sample of the images used to investigate the statistical properties of the portal imaging. Above: images of the head (left) and chest (right). Below: images of the pelvis in anterior-posterior projection (left) and lateral (right).



(a) Uniform image (b) Noise image

Fig. 2. Uniform image and noise image

the conservation of energy in the transform domain, which is necessary for the comparison of image and noise histograms.

2.4 Denoising Methods and Assesment

Wavelet thresholding takes advantage of the different statistical features for noise and signal. In the present work wavelet thresholding is compared to Wiener

filtering. The portal image in figure 3a (image x) is used to carry out this comparison. The procedure follows three steps: 1) Addition of white Gaussian noise n to x to obtain a noisy image $y = x + n$ (see figure 3b). 2) Denoising by means of one of the methods to obtain the estimation $\hat{x}$. 3) Evaluation of the denoised image.

Noises of different amplitudes are used in step 1 ranging from 25 dB to 35 dB of peak signal to noise ratio (PSNR). PSNR is defined as

$$PSNR = 10 \log_{10} \frac{Range^2/N}{MSE} \tag{1}$$

where Range is the difference between the maximum and the minimum in the image and MSE is the mean square error

$$MSE = \frac{1}{N}||n||^2 \tag{2}$$

On the other hand, the evaluation of the denoised image is carried out by means of the PSNR again (in this case $MSE = \frac{1}{N}||x-\hat{x}||^2$) and the structural similarity index SSMI

$$SSIM(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \tag{3}$$

where μ_x is the pixel mean value of image x representing luminance, σ_x is the standard deviation of x representing contrast, and $\sigma_{x\hat{x}}$ is the correlation coefficient between x and $\hat{x}$ that measures the structural similarity between both images. Constants C_1 y C_2 are small positive values to avoid denominator singularities.

(a) Pelvis AP (b) Pelvis AP plus noise

Fig. 3. A portal image of the pelvis before and after adding white Gaussian noise of 25 dB PSNR

The Wiener filter is an adaptive filter that estimates the local statistics of the image. In an neighbourhood of 3x3 pixels the mean pixel value μ and the variance σ^2 are calculated, the noise variance ν^2 is also estimated as the mean of the local variances. The pixel value is then calculated as

$$\hat{x}(i,j) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2} \left(x(i,j) - \mu \right). \tag{4}$$

Wavelet thresholding operates on the coefficients of the noisy image so that those coefficients with an amplitude under a given threshold are converted to zero. After that the resulting wavelet subbands are back-converted to the estimated (denoised) image. The technique was first implemented in 1991 [7], and was developed further by Donoho and Johnstone [8,9,10]. Donoho and Johnstone showed that denoising schemes by means of wavelet thresholding provide results close to optimal solution under the minimax criterion.

Thresholding is carried out by two different ways, hard (equation 5) and soft (equation 6) thresholding. The hard thresholding function is not continuous, resulting in an increase in the variance of the estimated coefficients which leads to artefacts in the estimated image. On the other hand, soft thresholding biases the transformed coefficients, over-smoothing the estimated image.

$$\theta_h(w) = \begin{cases} 0 & if\ |w| \leq\ T, \\ w & if\ |w| >\ T \end{cases} \tag{5}$$

$$\theta_s(w) = \begin{cases} 0 & if\ |w| \leq\ T, \\ sign(w)\,(|w| - T) & if\ |w| >\ T \end{cases} \tag{6}$$

A key element in thresholding is the value used as threshold. When noise is additive, Gaussian and white and a decimated orthonormal transformation is used Donoho and Johnstone proposed [3] the universal threshold

$$T_u = \hat{\sigma_n} \sqrt{2 \log(N)} \tag{7}$$

where $\hat{\sigma_n}$ is the estimation of the noise standard deviation and N is the number of the coefficients in a detail subband. This threshold is optimal under the minimax criterion but, in order to minimize the mean square error the Stein's unbiased risk estimator (SURE) threshold obtains better results. The SURE is an efficient risk or MSE estimator for the estimated image when the noise is additive and Gaussian [11]. Therefore, the minimization or SURE optimizes the PSNR.

The wavelet function used to transform the image is another decisive factor. The smoothness, symmetry, number of vanishing moments and support of the wavelet influence the efficiency of the processing and the visual quality of the estimated image. For this reason Symmlets (sym) are widely used wavelet functions in image processing. They have a minimal support for a given number of vanishing moments, and they are quasi symmetric and smooth. In this work the sym8, with eight vanishing moments, are the wavelets used in a decimated wavelet thansform. Also, soft thresholding and SURE threshold are used to carry out the thresholding procedure.

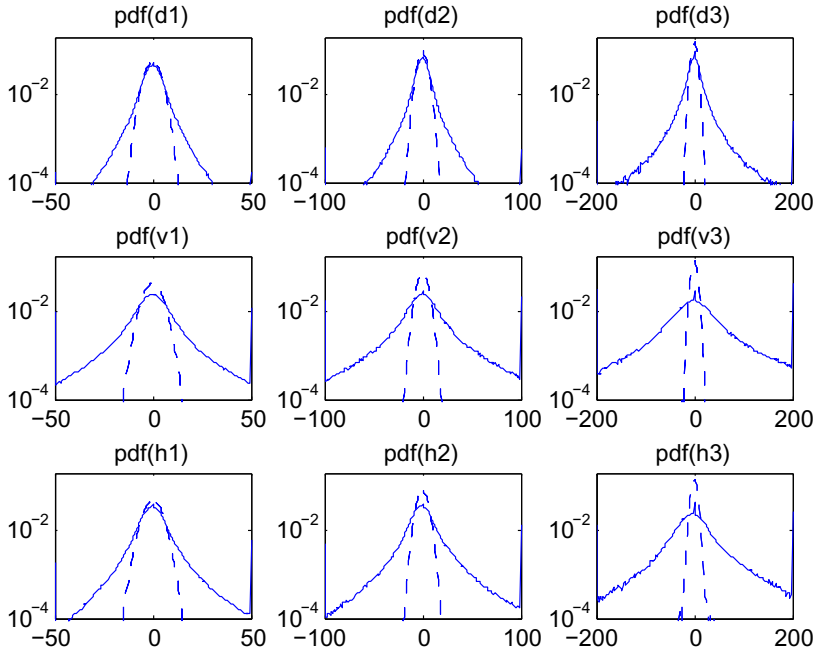


Fig. 4. Wavelet coefficients marginal pdf's for noise (dashed) and image (solid lines). The first three scales of orientations horizontal (h1-h3), vertical (v1-v3) and diagonal (d1-d3) are shown.

3 Results and Discussion

The comparison between the marginal pdf's of noise and image, in the wavelet domain, is shown in figure 4. This figure represents, on a logarithmic scale, the curves for the experimental pdf's of portal (solid) and noise (dashed line) images. The coefficients in the first three scales for diagonal (d1-d3), vertical (v1-v3) and horizontal (h1-h3) orientations are represented. A decrease in the relative importance of noise with increasing scale can be observed. The amplitude of the noise distribution reduces with respect to the amplitude of the image distribution as the scale grows. The curves for the noise show a parabolic shape across the different scales, which translates into a Gaussian character in the transformed domain. On the other hand, the pdf's for portal images show longer tails than expected for a Gaussian distribution.

The shape of the pdf's in the wavelet domain for noise and image suggest thresholding as an efficient approach to noise reduction. Removing small amplitude coefficients in the transformed noisy image results in a much higher reduction in the noise power than in the signal power.

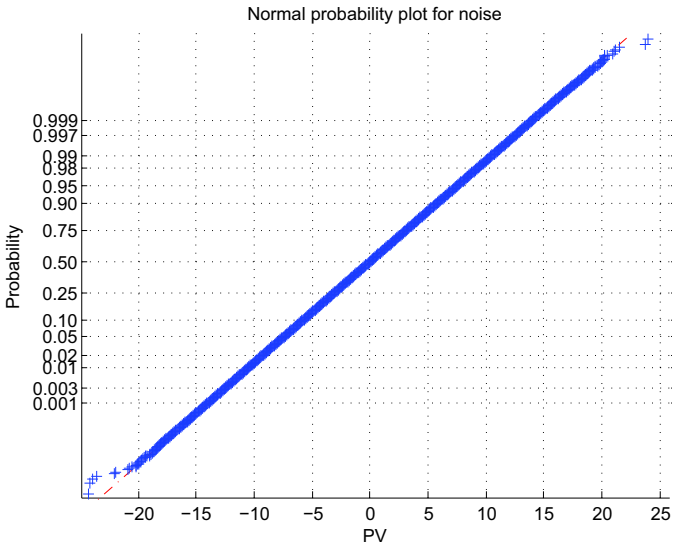
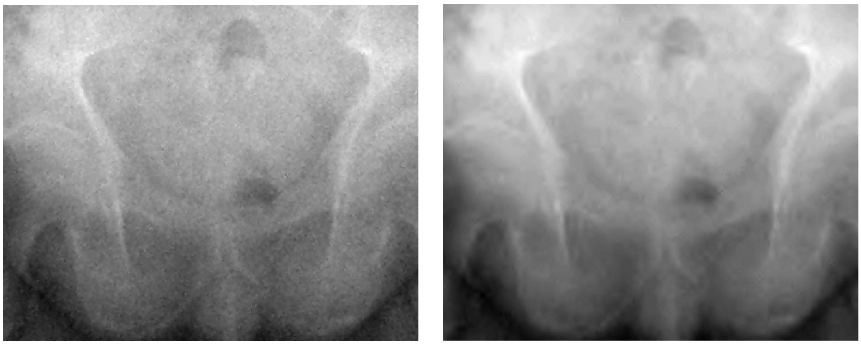


Fig. 5. Normal probability plot for noise values in the image domain

The distribution of noise in the image domain is clearly Gaussian, as shown in figure 5. The Gaussian shape of the noise pdf justify the selection of the SURE threshold.

Figure 6 shows the results of applying the Wiener and the thresholding methods to the noisy image in figure 3b.



(a) Wiener (b) Wavelet thresholding

Fig. 6. Noise reduction of the image in figure 3b

The PSNR and SSIM values for the estimated images are shown in tables 1 and 2. These tables show the results for the three levels of added noise. Despite being quite different, the PSNR and SSIM values show similar differences between both denoising methods and a similar behaviour with increasing noise. Thresholding always gives a better performance than Wiener filtering and, as the noise PSNR rises, the differences between both methods increase.

Table 1. PSNR (in dB) for the estimated images of the noisy pelvic image

noise PSNR	Wiener	Wavelet thresholding
25	32.4	38.0
30	37.0	40.3
35	40.9	42.3

Table 2. SSIM for the estimated images of the noisy pelvic image

noise PSNR	Wiener	Wavelet thresholding
25	0.735	0.928
30	0.883	0.947
35	0.947	0.961

4 Conclusions

Noise in portal imaging systems resembles Gaussian noise. The Gaussian nature of noise distribution justifies the use of noise reduction methods based on minimizing the Stein's unbiased estimator (SURE), such as the selection of the SURE threshold in wavelet thresholding methods.

Noise distribution remains Gaussian in the transform domain, and this fact represents the first major difference with the case of portal images (see Figure 4). For portal images the distribution of the coefficients in the wavelet domain has larger tails than for the noise and is sharper in the vicinity of the distribution centre. The importance of the detail coefficients of large amplitude in the image distribution is due to irregularities in the image such as corners, edges or spikes. The differences found between the distributions of portal images and noise in the portal imaging system is similar to those found between natural images and Gaussian white noise. Therefore the noise reduction methods based on the thresholding of wavelet coefficients, very efficient in the case of natural images, are also effective in the case of portal images.

Thresholding the wavelet coefficients of the noisy image gives better denoising results than the Wiener filter in all the analysed cases. The differences between the methods increase as the noise increases. Both metrics, PSNR and SSIM, gave basically the same results.

References

1. Herman, M., Balter, J., Jaffray, D., McGee, K., Munro, P., Shalev, S., Van Herk, M., Wong, J.: Clinical use of electronic portal imaging: Report of AAPM Radiation Therapy Committee Task Group 58. Med. Phy. 28(5), 712–737 (2001)
2. Antonuk, L.: Electronic portal imaging devices: a review and historical perspective of contemporary technologies and research. Phys. Med. Biol. 47(6), R31 (2002)
3. Donoho, D., Johnstone, J.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3), 425–455 (1994)
4. Simoncelli, E., Adelson, E.: Noise Removal Via Bayesian Wavelet Coring (1996)
5. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process 12(11), 1338–1351 (2003)
6. Ferrari, R.J., Winsor, R.: Digital radiographic image denoising via wavelet-based hidden Markov model estimation. Journal of Digital Imaging 18(2), 154–167 (2005)
7. Weaver, J., Yansun, X., Healy, D., Cromwell, L.: Filtering noise from images with wavelet transforms. Magnetic Resonance in Medicine 21(2), 288–295 (1991)
8. Donoho, D.: De-noising by soft-thresholding. IEEE Transactions on Information Theory 41(3), 613–627 (1995)
9. Donoho, D., Johnstone, J.: Adapting to Unknown Smoothness via Wavelet Shrinkage. Journal of the American Statistical Association 90, 1200–1224 (1995)
10. Vidakovic, B.: Nonlinear wavelet shrinkage with bayes rules and bayes factors. J. of the American Statistical Association 93, 173–179 (1998)
11. Stein, C.: Estimation of the Mean of a Multivariate Normal Distribution. The Annals of Statistics 9(6), 1135–1151 (1981)

Self Organized Biogeography Algorithm for Clustering

Leila Hamdad[1], Anissa Achab[2], Amira Boutouchent[2], Fodil Dahamni[2], and Karima Benatchba[2]

[1] Ecole Nationale Supérieure d'Informatique, BP 68M, OuedSmar, 16307, Alger, Algrie, LCSI
l_hamdad@esi.dz

[2] Ecole nationale Supérieure d'Informatique, BP 68M, OuedSmar, 16307, Alger, Algrie, LMCS
{a_achab,a_boutouchent,f_dahamni,k_benatchba}@esi.dz

Abstract. We propose in this work a new self organized biomimitic approach for unsupervised classification, named BFC, based on BBO (Biogeography based optimization). This method is tested on several real datasets(IRIS, Satimages and heart). These benchmarks are characterized by increasing overlap degree. Moreover, a comparison of BFC with other clustering methods having proven their efficiency is presented. We will highlight the impact of this overlap on the performance of the methods.

Keywords: Clustering, Self organization, Biomimetic method, Biogeography, Biogeography based optimization.

1 Introduction

This paper deals with one of the main task of data-mining which is clustering. Formally, it consists on partitioning a set S of N objects, $S = \{O_1, O_2, \ldots O_N\}$, based on a similarity metric, into a number of clusters $(C_1, C_2, \ldots\ C_K)$, such as

$$\begin{cases} C_i \neq \emptyset\ for\ 1 \leq i \leq K, \\ C_i \cap C_j = \emptyset, 1 \leq i, j \leq K, i \neq j \\ S = \cup_{i=1}^{K} C_i. \end{cases}$$

Objects of a cluster must be similar, while objects of different clusters must be dissimilar. It is an important task in data mining as it enables to show interesting objects grouping without a priori knowledge.

Clustering has been shown **NP**_difficult for several space metrics [2] and it is in this context, that the use of metaheuristics, seems very appropriate. There has been a growing interest in biomimetic metaheuristics to solve this problem as they have given good results for many problems. Their strength comes from their interesting use of two strategies during their search process: intensification

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 396–405, 2013.

and diversification. Among these methods, one can find genetic algorithms [1], ant colonies [3], particle swarm [4], and cellular automata [5],[6].

In this paper, we are interested in a more recent bio inspired metaheuristic named biogeography based optimization (BBO), to solve clustering. It is a self-organized method based on biogeography, introduced by Simon in 2008 [7]. It aims to optimize species' distribution in an ecosystem through migration process. The ecosystem is governed by its own simple rules and is completely autonomous[7]. Different species migrate adaptively to different habitats. The species migration process from one habitat to another leads the ecosystem to evolve dynamically to an optimized distribution.

This paper is organized as follows, in the next section, we will present the biogeography based optimization method (BBO). A new self organized unsupervised clustering method, named BFC, based on BBO is presented in section 3. The performance of this algorithm is tested on different benchmarks, Iris, Heart, Glass and Satimages [1]. Results of these tests are given in section 4. We conclude this section with a comparison between BFC and another self organized method (Automata Cellular).

2 Biogeography Based Optimization

Biogeography-based optimization (BBO) is based on the study of geographical distribution of biological organisms (biogeography) [7]. It uses mathematical models to describe the species migration among different islands, and species extinction or emergence [7]. This method can be classified as an evolutionary population based metaheuristic. It has been used to solve several optimizations problems such as image segmentation [9], block matching [10] and constrained optimization problems [11].

2.1 Natural Inspiration

Biogeography is a natural way (method) for species (Penguins, Penguins, ...) distribution. Indeed, in each island, lives a number of biological species. If an island is considered "good", it will attract more species and therefore have large populations (high immigration rate)[7]. As a consequence, its resources will tend to decline (reduce), and the species will tend to migrate to other islands. Migration process generates movement of species among islands. In an ecosystem (set of islands), the habitability of an island (HSI) is characterized by a set of variables named quality suitable variables (SIV). These variables can be food, climate, vegetation.... They are used to explain the dependent variable (HSI)[7]. HSI is proportional to biological diversity of a habitat. Hence for, habitats with high HSI will tend to attract a big variety of species and end up having a large population. As a result, these habitats will tend to have a low immigration rate as

they become saturated. But, nearby habitats will have a high emigration rate[7]. On the other hand, a habitat with a low HSI will tend to have a less diversified number of species and a high immigration rate. If species immigrate to these habitats their HSI rate will grow. However, if no immigration occurs then the species of these habitats may extinct.

As different species migrate from one habitat to another, each habitat (Hi) is characterized by an immigration rate (λ_i) and emigration rate (μ_i) defined respectively below as:

$$\lambda_i = I(1 - \frac{i}{n}), \quad i = 1, ..., n.$$

And

$$u_i = \frac{E_i}{n}, \quad i = 1, ..., n.$$

Where n represent the maximal number of species in the habitat, E is the maximum emigration rate and I is the maximum possible immigration rate to the habitat .

2.2 Artificial Biogeography Process

The biogeography process presented above can be modeled to solve optimization problems. Each habitat will represent a solution to the problem which can be represented as a vector of the problem characteristics. A solution's quality can be evaluated using an objective function (HSI). A solution is considered as a good one if it has a high HSI [12]. Immigration (λ_i) and emigration (μ_i) rates are associated to each solution and are used to modify SIV values, thus producing new solutions. With a certain probability, characteristics of a solution will be replaced by characteristics of the other ones. In BBO, the migration and mutation processes are defined as follows:

• **Migration:** It is the adaptive process that is used to modify existing solutions (habitats) [7]. Suppose that a given Solution (S_i) is selected to be modified. Its immigration rate λ_i is used to decide whether each of its quality index variables (SIV) will be modified or not. If a given SIV in S_i is selected for modification, the emigration rates μ_i of the other solutions are used to decide probabilistically which solution (S_j) will migrates one of its SIV, chosen randomly, to solution S_i. The HSI of solution S_i will be modified accordingly.

• **Mutation:** Cataclysms can occur in natural habitats, modifying completely its HSV. This phenomenon is reproduced in the BBO. It consists on a mutation that will change some of the solution characteristics according to a probability.

3 Biogeography for Clustering(BFC)

In this work, we propose an algorithm (BFC) for clustering based on BBO. We consider a set of habitats (ecosystem) which represent clusters. Initially,

the number of habitats in the ecosystem is set to K. It does not represent the final number of clusters. In fact, when the stopping criterion is reached, habitats with a low number of objects will be deleted from the ecosystem (cataclysm phenomenon). As a result, their objects will not be classified.

The algorithm starts by storing the objects to be classified in a universal habitat (UH) [13][13]. Then, similarity threshold *simt* is set to the smallest Euclidian distance among all objects of the dataset. The algorithm's main step is the migration process which is repeated until stopping criteria is reached. This process begins by objects' migration (immigration/emigration) between UH and all habitats $(H_i, 1 \leq i \leq K)$ (**Algorithm 1**), then migration among habitats (**Algorithm2**). During the first migration, UH's objects immigrate to most adapted habitats if they exist. A habitat (H_i) is considered most adapted to object O_j if H_i's mean similarity including O_j is less than *simt*. During the second migration, objects from different habitats will try to find a better suited one to immigrate to. After completion of a migration process, when no more objects moves are possible, *simt* is increased as follows:

$$simt = simt + \frac{Standard\ deviation\ of\ the\ similarity\ matrix}{C},$$

where C is a constant.

Algorithm 1. Migration of objects between UH and H_i, $(1 \leq i \leq K)$

Begin

1. For each object (O_j) of the UH,
2. O_j immigrates to H_i If the mean similarity of H_i including O_j is less than *simt*

End

Algorithm 2. Migration process among Habitats

Begin

1. For each habitat (H_i) of the ecosystem,
2. For each object (O_j) of H_i,
3. O_j immigrates to H_k If the mean similarity of H_k including O_j is less than *simt*.

End

This migration process is repeated until a stability is reached (no movement is possible among habitats (no immigration or emigration) or a stagnation (same movements are repeated for a certain number of iterations).

The proposed biogeography for clustering algorithm is given in **Algorithm3**.

Algorithm 3. BFC

Let $\{\mathbf{o}_1, ..., o_n\}$ a set of n data to classify and K, the number of initial classes.
Begin

1. Create a universal habitat (UH).
2. Create a similarity matrix of data in universal habitat to define *simt*
3. Create an ecosystem of K habitats.
4. *simt* = maximal similarity between the set of data
5. Repeat
   a. Migration process for objects between UH and Hi $(1 \leq i \leq K)$
   b. Migration of objects among habitats Hi $(1 \leq i \leq K)$
6. Increase *simt*.
7. until stopping criterion is reached
8. Delete the Habitats with less than l objects after having returned those objects to the UH.

End

4 Tests and Results

To test the performance of BFC, we conducted several tests on different supervised benchmarks, Iris, Heart, and satimages [8]. Their respective PCA representation is given in Fig.1. The first dataset (Iris) has three classes with 50 objects each. One class is linearly separable from the other two. The two others overlap as they have similar objects. The dataset Heart has 270 objects with 14 features. It consists of two classes of respectively 150 and 120 objects. This benchmark is difficult as the two classes overlap. The last dataset has 2000 objects having each 37 features. It consists on 6 classes of 460, 223, 395, 210, 236 and 476 objects. Satimages is a difficult benchmark because several classes overlap.

As these benchmarks are supervised, the reliability measure (F)we used to evaluate the performance of our algorithm is given in formula 1. Let o_i and o_j, two subjects to classify, and $c(o_i)$ the original class of the subject o_i and $C_1(o_i)$, the class assigned by the algorithm. Reliability F is defined by:

$$F = 1 - E \tag{1}$$

where

$$E = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \varepsilon_{ij}$$

with

$$\varepsilon_{ij} = \begin{cases} 0 & if \;\; (c(o_i) = c(o_j) \;\; and \;\; C_1(o_i) = C_1(o_j)) \;\; or \;\; (c(o_i) \neq c(o_j) \;\; and \;\; C_1(o_i) \neq C_1(o_j)) \\ 1 & else. \end{cases}$$

Where $c_k(o_i)$ and $C_k(o_i)$ are respectively the obtained and the original class for object o_i.

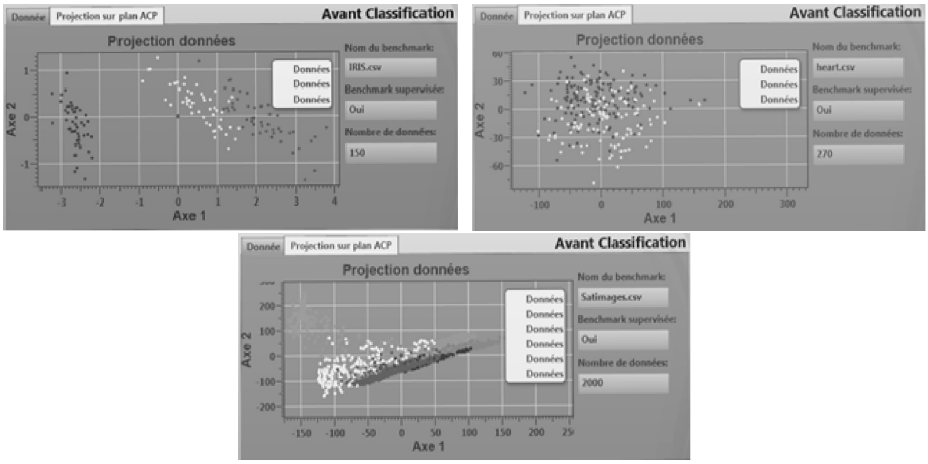


Fig. 1. PCA representation of datasets Iris, Heart and Satimage

For all results given below, we executed the algorithm with the same set of parameters 10 times.

4.1 Tests on Benchmark IRIS

For this benchmark, we initialized the number of habitats in the ecosystem to 3, and *simt* to 0.1. The following graph shows the evolution of the best (BR), worse (WR) and average reliability (AR) according to *simt*. We note that as *simt* increases, BR, WR and AR increase too. Their best value, respectively 88.36%, 79.45% and 84.44%, are reached for $simt = 1$. The rate of unclassified objects is 2.67%. Above this value of *simt*, all objects are classified, but the

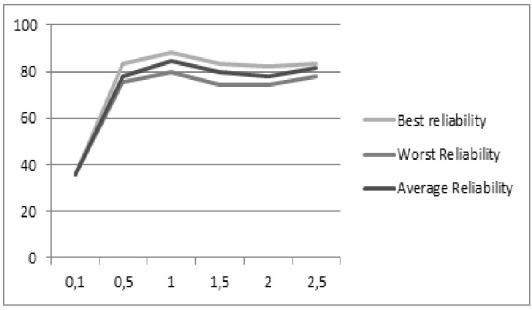


Fig. 2. Evolution of BR, WR, and AR according to Simt

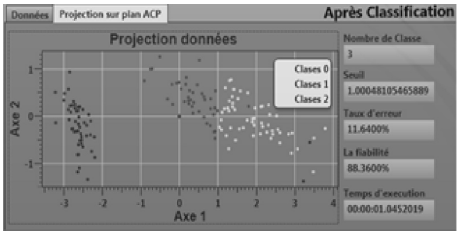


Fig. 3. Clustering of dataset Iris with BFC

reliability decreases. This can be explained by the migration process among habitats. As *simt* increases, accepting an object in a class is less stringent. As a result, an object can easily move from one habitat to another (case of overlapping habitats). The best value for BR, 88.36%, is obtained because two classes overlap (see Fig.2). As result, some objects of these classes are misclassified as they are similar to both classes. Indeed, those two classes have respectively 90% and 75.08% of objects well classified. The objects of the remaining class, linearly separable from the other two, are well classified (100%). All objects are classified (Fig.3).

4.2 Tests on Benchmark Satimages

We initialized *simt* to 13.50 and the number of introduced clusters to 6. Figure Fig.4 shows the evolution of BR, AR and WR according to *simt* for the benchmark Satimages. We note that the three curves are very close. The maximum reliability obtained is 81.53% for a threshold equal to 50 (Fig.4). Fig.5 represent BFC clustering result on Satimages.

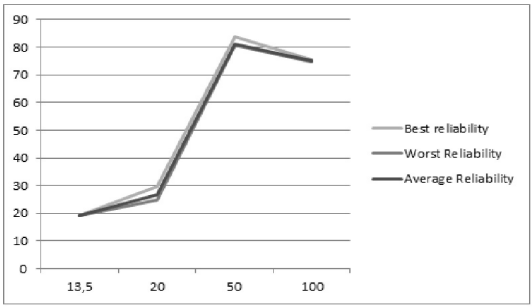


Fig. 4. Evolution of BR, WR, and AR according to Simt

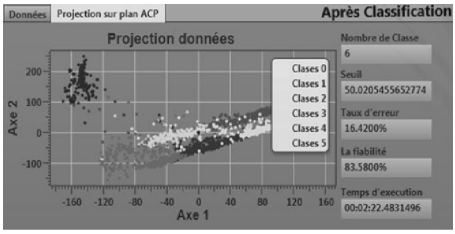


Fig. 5. Clustering of dataset Setimage with BFC

4.3 Comparison of BFC with Other Methods

We compare BFC to Kmeans, PSO, Hybridization of PSO and K-means and cellular automata on the three benchmarks Iris, satimages and heart. The results are presented in the Fig.7. We notice that for benchmark Iris, characterized by a small overlapping degree, the best clustering is given by PSO (92,49%) followed by BFC (88,36%). Both methods outperform K-Means (88%). For benchmark Satimages, PSO becomes less efficient as the degree of overlapping is more important. The best result is reached with K-Means (85%) followed by BFC (81%).

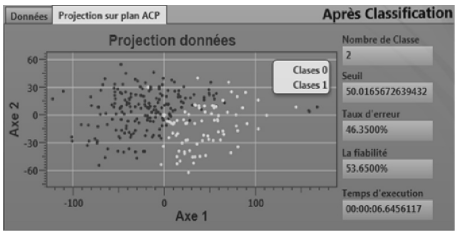


Fig. 6. Clustering of dataset Heart with BFC

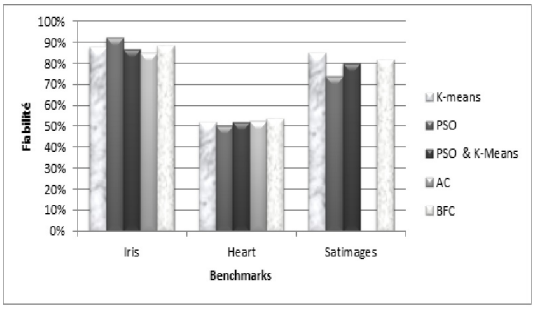


Fig. 7. Benchmarks

When tested on a high overlapping degree benchmark, Heart, BFC outperforms all tested methods. This is due to the fact that during last iterations the solution is improved as objects migrate from one habitat to another more adapted one. Fig.6 gives BFC clustering of dataset heart.

5 Conclusion

In this paper, we dealt with an important task in data mining which is clustering. It has been classified as an NP-difficult problem. We presented a method based on biogeography we named BFC. We tested this method on three real datasets with different overlapping degrees. Moreover, we compared it to four different types of methods: K-means, PSO, hybridization of PSO and K-means, and cellular automata. Eventhough BFC was outperformed by PSO and K-means on respectively Iris and Satimage, it showed its efficiency when used on heart, a dataset with high overlap degree.

References

1. Maulik, U., Bandyopadhyay, S.: Genetic algorithm-based clustering technique. Pattern Recognition 33, 1455–1465 (2000)
2. Ostrovsky, R., Rabani, Y.: Polynomial time approximation schemes for geometric k-clustering. In: IEEE Symposium on Foundations of Computer Science, pp. 349–358 (2000)
3. Chen, L., Xu, X., Chen, Y.: An adaptative ant colony clustering algorithm. In: Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, August 26-29 (2004)
4. Chen, C.Y., Ye, F.: Particle Swarm Optimization Algorithm and Its Application to Clustering Analysis. In: Proceedings of the IEEE ICNSC, Taipei, Taiwan, pp. 789–794 (2004)
5. Azzag, H., Picarougne, F., Guinot, C., Venturini, G.: Classification de données par automate cellulaire. Comptes rendus des 12-émes Rencontres de la Société Francophone de Classification, Mais-1er Juin, Université du Québec á, 30 (2005)
6. Kari, J.: Theory of cellular automata: A survey. Theoretical Computer Science 334(1-3), 3–33 (2005)
7. Simon, D.: Biogeography-Based Optimization. IEEE Transactions on Evolutionary Computation 12(6) (December 2008)
8. Blake, C., Merz, C.: CI Repository of machine learning databases. University of California Irvine, Dept. of Information and Computer Sciences (1998), `http://www.ics.uci.edu/mlearn/MLRepository.html`
9. Chatterjee, A., Siarry, P., Nakib, A., Blanc, R.: An improved biogeography based optimization approach for segmentation of human head CT-scan images employing fuzzy entropy. Engineering Applications of Artificial Intelligence 25(8), 1698–1709 (2012)
10. Zhang, P., Wei, P., Yu, H.Y.: Biogeography-based optimisation search algorithm for block matching motion estimation. IET Image Processing 6(7), 1014–1023 (2012)

11. Boussaid, I., Chatterjee, A., Siarry, P., Ahmed-Nacer, M.: Biogeography-based optimization for constrained optimization problems. Computers and Operations Research 39(12), 3293–3304 (2012)
12. Wesche, T., Goertler, G., Hubert, W.: Modified habitat suitability index model for brown trout in southeastern Wyoming. North Amer. J. Fisheries Manage. 7, 232–237 (1987)
13. Panchal, V.K., Singh, P., Kaur, N., Kundra, H.: Biogeography based Satellite Image Classification. International Journal of Computer Science and Information Security (IJCSIS) 6(2), 269–274 (2009)

Automatic Orientation of Functional Brain Images for Multiplataform Software

I. Alvarez Illán[1], Juan Manuel Górriz[1], Javier Ramirez[1], Diego Salas-González[1], Francisco Jesús Martínez-Murcia[1], F. Segovia[3], and C.G. Puntonet[2]

[1] Dept. of Signal Theory, Networking and Communications
University of Granada, Spain
[2] Dept. of Computers Architecture and Technology
University of Granada, Spain

Abstract. An automated method for orientation of functional brain image is proposed. Intrinsec information is captured from the image in three stages: first the volume to identify the anterior to posterior line, second the symmetry to detect the hemisphere dividing plane and third the contour to determine the up-down and front-back orientation. The approach is tested in more than a tousand images from different formats and modalities with high reconition rates.

1 Introduction

There exist a wide variety of modalities for brain imaging archives. The most extended format used in major manufacturers of medical imaging equipment (e.g., GE, Siemens, Philips) is the DICOM (Digital Imag-ing and Communications in Medicine; http://medical.nema.org/). The DICOM standard provides interoperability across hardware, but was not designed to facilitate ecient data manipulation and image processing. Hence, additiona data formats have been developed over the years to accommodate data analysis and image processing. For instance, the ANALYZE format or the NIfTI format, developed in an effort to improve upon the DICOM format. The compatibility between these formats depends on the ability to transform the information encoded in the headers. The spatial orientation is codified in several different ways (linear tranformations, angles, origins, etc...) depending on the format. In a multiplataform software, supporting several brain imaging formats, this lack of homogeneity may contitute a drawback. Moreover, the process of anonimizing the images, necessary for research studies, may also damage the orientation information in the header of a brain image file, corrupting the information for recovering the brain orientation.

The problem of automatically orient a brain image has been faced several times in the literature (examples are [1, 2]). There exist two main motivations to study this problem. The first is the identification of the mid-sagittal plane of the brain in magnetic resonance images (MRI) for medical reasons. There exist a vast literature concerning the study of this particular problem (for example

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 406–411, 2013.

[3, 4, 5, 6, 7]) with techniques that sometimes can be extended to functional brain images. The second is for registration propose to a common space for comparisons [8, 2].

In this work we study the problem of automatically orient brain images in the context of functional images, where anatomical details are not as well defined as in MRI images. Our aim is not the registration of the images nor to identify only the MSP, but to manage and store brain image information in a multiplataform software. This first step may also serve for posterior registration porpuses, as many spatial information from the brain would be extracted.

2 Methods

For automatically orient an image, a combination of three different techniques is employed; morphological transformations, cross-correlation and third order statistical moments. First, morphological transformations are applied for the detection of the brain surface. Once the surface is extracted, its corresponding ellipsoid is calculated. The mayor axis of the ellipsoid is identified with the direction of the anterior comissure to posterior comissure line (AC-PC). Secondly, the cross-correlation between the image and reflected versions of it, is calculated in order to identify the mid-sagittal plane. Third, the orientation in the remaining directions is identified by extracting parameters of the image based on statistical properties.

2.1 Morphological Operations

For identifying the brain surface, the image is binarized only considering intensities above a threshold. This value is fixed using Otsu's method[9], which chooses the threshold to minimize the intraclass variance of the black and white pixels. A slice-by-slice morphological opening is applied to the resulting binarized image, that is, an erosion operation followed by a dilation operation using a predefined disk-sized structure. Also, inside holes are filled.

The resulting image is a volume image in which small external artifacts have been removed. From the moments of this volumen image, the principal axis of the corresponding ellipsoid are calculated, and the anterior to posterior line is identified with the mayor axis.

2.2 Cross Correlation

An important characteristic of the brain morphology that can be used to identify the orientation is the mid-sagittal plane (MSP) symmetry. To determine the symmetry of the brain image, [4] and [8] compare it with the image that is obtained when it is reflected (flipped) about the expected MSP plane. The measure of symmetry used is the cross-correlation between the original and the fliped image. Based on these old the ideas, we use here cross correlation to identify the saggital plane.

To achieve the goal, we first project the whole image in the AC-PC direction, creating a 2D image $f(x, y)$. This image still preserves the MSP symmetry. From the above obtained surface, we extract a window of the image that contains only the brain g and flip it. The flip is made in 1) the up-down direction (g_1) or 2) in the left-right direction (g_2). We compute the normalized cross correlation between f and g_i as:

$$nxc_i = \frac{1}{n} \sum_{x,y} \frac{(f(x,y) - \overline{f})(g_i(x,y) - \overline{g_i})}{\sigma_f \sigma_{g_i}} \tag{1}$$

and look for a maximum in nxc_i, where n is the number of pixels in $g_i(x, y)$ and $f(x, y)$, $\overline{f}$ is the average of f and σ_f is the standard deviation of f. The maximum value between $max(nxc_1)$ and $max(nxc_2)$ will determine the direction perpendicular to de MSP.

2.3 Statistical Parameters

Once the MSP is determined, the volume image defined above is projected ortogonally into two lines: a) the line defined by the intersection between the axial and the sagittal planes, and b) the line defined by the intersection of the coronal and sagittal planes. These two lines would represent the contour of the image in these directions.

Interpreting these curves as probability distributions, it is possible to calculate the skewness of the distribution that would give a measure of the asymmetry of the curve, defined as:

$$skew = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^3}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2\right)^{3/2}} , \tag{2}$$

where x_i are the points in the curve, $\overline{x}$ is the mean, and n the number of points in the curve. A positive value of the skewness would indicate that values of the curve are accumulated asymmetrically towards the left, and a negative value would indicate the opposite. The sign of the skewness of the projected image in the lines would help us to determine de anterior to posterior orientation, togheter with the up-down orientation.

2.4 Databases

The algorithm is tested in more than 1000 images from different types of functional brain images. SPECT and PET images are tested, and also different radiotracers as ^{99m}Tc-ECD, ^{18}F-FDG, and FP-CIT-I-123 (DaTSCAN).

- **Database 1:** The database consists of a set of 97 3D SPECT brain images produced with an injected gamma emitting ^{99m}Tc-ECD radiopharmeceutical and acquired by a three-head gamma camera Picker Prism 3000. Images of the brain cross sections are reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter.

- **Database 2:** A total of 208 images conform this subset. The images were obtained between 3 and 4 hours after the intravenous injection of 185 MBq (5 mCi) of Ioflupane-I-123, with prior thyroid blocking with Lugol's solution. The tomographic study (SPECT) with Ioflupane/FP-CIT-I-123 was performed using a General Electric gamma camera, Millennium model, equipped with a dual head and general purpose collimator. A 360-degree circular orbit was made around the cranium, at 3-degree intervals, acquiring 60 images per detector with a duration of 35 seconds per interval, each consisting of a 128 $\times$ 128 matrix. Transaxial image slices were reconstructed using the filtered back-projection algorithm without attenuation correction, and applying a Hanning filter (cutoff frequency equal to 0.7).
- **Database 3 and 4:** 403 PET images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) are collected in database 3, together with 301 SPECT images from the Parkinson's Progression Markers Initiative (PPMI) of database 4. All of them are spatially pre-processed, and spatially aligned in the AC-PC line.

3 Results and Discussion

Tables 1 and 2 summarize the recognition rates in the different planes. Table 1 is calculated using only normalized images, while 2 uses only raw data. The columns in the tables correspond to the different recognition steps, in which each of the possible orientations are detected:

- AC-PC line: is the detection of the AC-PC line by identifying it with the mayor axis of the ellipsoid.
- MSP : is the detection of the symmetry plane by cross correlation computation.
- Ant-Post: is the detection of the orientation in the Anterior-to-posterior direction.
- Up-Down: is the detection of the orientation in the coronal plane.

Table 1. Percent of correctly indentifications with normalized images

	AC-PC line	MSP	Ant-Post	Up-down
Database 1	100.00 %	100.00 %	100.00 %	100.00 %
Database 2	100.00 %	100.00 %	98.65 %	100.00 %
Database 3	100.00 %	100.00 %	99.00 %	99.67 %
Database 4	100.00 %	100.00 %	100.00 %	100.00 %

From the first column of tables 1 and 2, the identification of the mayor axis of the ellipsoid with the AC-PC line is completely justified. The second column

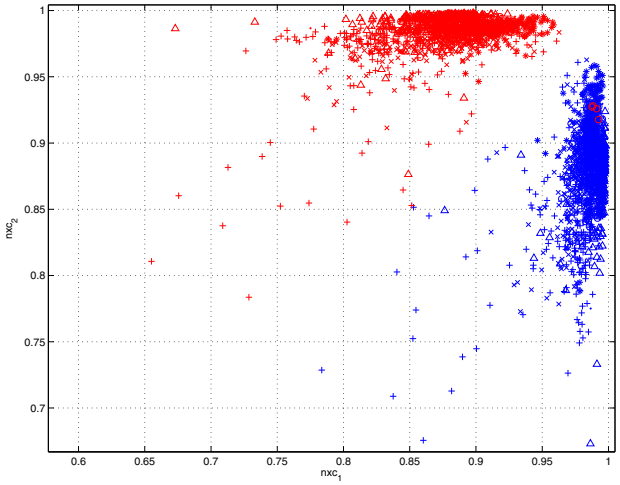


Fig. 1. Normalized cross correlation of f and g_i (blue) and for g_i and a $\pi/2$ rotation of f (red): * Database1, x Database2, o Database3, △ Database 4, · Database1 (raw), x Database2 (raw)

in both tables also support the use of cross correlation to identify the MSP. The details of the values of eq. 1 are plotted on figure 1, where the complete separation between nxc_1 and nxc_2 is made visible.

The remaining columns report differences in the recognition rate depending on the anatomical details present in the images. In the case of images with low anatomical details obtained with DaTSCAN SPECT, the algorithm has some miscalssification results. It is to be noted here that databases 2 and 3 contain Parkinson's disease affected images, in which severe dopaminergic deficits make the task of orientation very difficult, even visually.

Table 2. Percent of correctly indentifications with raw data

	AC-PC line	MSP	Ant-Post	Up-down
Database 1	100.00 %	100.00 %	100.00 %	92.78 %
Database 2	100.00 %	100.00 %	96.19 %	94.23 %

Also differences are found in the algorithm behaviour depending on the preprocessing of the images. These differences only come in the up-down and anterior-posterior orientations, giving a slightly less accurate result in the case of raw data, as it could be expected. This results could be improved by looking for alternative parameters to skewness, as skewness may not respresent correctly the asymetry in all the cases for a finite sample[10].

4 Conclusions

We have presented a robust and efficient algorithm for automatic recognition of brain image orientation. The method is robust against the modality of brain image and has remarkable success in raw data. Automating this task has important applications in multiplataform brain imaging software and registration techinques.

Acknowledgment. This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) and HD2008-0029 projects and the Consejera de Innovación, Ciencia y Empresa (Junta de Andaluca, Spain) under the Excellence Project TIC-02566.

References

[1] Christensen, J.D., Hutchins, G.C., McDonald, C.J.: Computer automated detection of head orientation for prevention of wrong-side treatment errors. In: AMIA Annual Symposium Proceedings, pp. 136–140 (2006) PMID: 17238318 PMCID: PMC1839503

[2] Tsao, J., Stundzia, A., Ichise, M.: Fully automated establishment of stereotaxic image orientation in six degrees of freedom for technetium-99m-ECD brain SPECT. Journal of nuclear medicine: official publication, Society of Nuclear Medicine 39(3), 503–508 (1998) PMID: 9529300

[3] Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3-d brain images. IEEE Transactions on Medical Imaging 21(2), 122–138 (2002)

[4] Ardekani, B., Kershaw, J., Braun, M., Kanuo, I.: Automatic detection of the midsagittal plane in 3-d brain images. IEEE Transactions on Medical Imaging 16(6) (December 1997)

[5] Hu, Q., Nowinski, W.L.: A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal. NeuroImage 20(4), 2153–2165 (2003)

[6] Ardekani, B.A., Bachman, A.H.: Model-based automatic detection of the anterior and posterior commissures on MRI scans. NeuroImage 46(3), 677–682 (2009) PMID: 19264138

[7] Liu, Y., Collins, R., Rothfus, W.: Robust midsagittal plane extraction from normal and pathological 3-d neuroradiology images. IEEE Transactions on Medical Imaging 20(3), 175–192 (2001)

[8] Junck, L., Moen, J.G., Hutchins, G.D., Brown, M.B., Kuhl, D.E.: Correlation methods for the centering, rotation, and alignment of functional brain images. Journal of Nuclear Medicine 31(7), 1220–1226 (1990)

[9] Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)

[10] von Hippel, P.T.: Mean, median, and skew: Correcting a textbook rule. Journal of Statistics Education 13(2) (2005)

Solving Number Series with Simple Recurrent Networks

Stefan Glüge and Andreas Wendemuth

Faculty of Electrical Engineering and Information Technology, Cognitive Systems Group, Otto von Guericke University Magdeburg and Center for Behavioral Brain Science, Universitätsplatz 2, 39106 Magdeburg, Germany
{stefan.gluege,andreas.wendemuth}@ovgu.de

Abstract. Number series tests are a popular task in intelligence tests to measure a person's ability of numerical reasoning. The function represented by a number series can be learned by artificial neural networks. In contrast to earlier research based on feedforward networks, we apply simple recurrent networks to the task of number series prediction. We systematically vary the number of input and hidden units in the networks to determine the optimal network configuration for the task. While feedforward networks could solve only 18 of 20 test series, a very small simple recurrent network could find a solution for all series. This underlines the importance of recurrence in such systems, which further is a basic concept in human cognition.

1 Introduction

Number series tests are popular at assessment centres and intelligence tests to measure a person's ability of numerical reasoning. Usually those tests utilize numerical sequences that are based on simple arithmetics. The underlying logical rule has to be deduced from an initial sequence. People are then asked to predict the next number according to the rule. Solving a number series requires deductive and inductive reasoning: deduction to find the logic between given numbers, and induction to predict the next number. The difficulty level of this task is controlled by two factors: first, the complexity of the logic behind the sequence, and second, the length of the temporal context the next element of the sequence depends on. For instance, the next element might not just depend on the last, but on the two last elements, etc. The cognitive load of these tasks is, for instance, discussed as early as 1986 by LeFevre and Bisanz [10].

In number series principally any mathematical function may be represented and the set of possible operators is not restricted. Therefore, not all series have the same difficulty. It is hard to define a general measure for the difficulty of a test series. The most obvious measure is the number of operations required to solve it. As this is a symbolic measure it most likely does not reflect the difficulty as it is perceived by humans.

Whereas there is a multitude of literature on automatic prediction of empirical time series (e.g. sunspot series [12], financial data [15], load forecasting [1]),

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 412–420, 2013.

few work has been devoted to solving arithmetically grounded number series, and to do so with small networks. So far, Ragni and Klein [14] made a first systematic attempt to apply artificial neural networks to the problem of number series prediction. In general, the function represented by a number series may be learned with a neural network as it is suited for regression problems [3]. However, in [14] Feedforward Neural Networks (FFNs) were applied. This network type is unable to learn an internal representation with a sequential character because of its inability to store its internal states for some period of time. Therefore, one needs the assumption that the sequence of relevant elements is of a fixed length. Typically, a set of input units for every event presented at time t to $t + n$, with n denoting the length of the time interval, is used [4]. This approach is often referred as *moving window* or *sliding window*.

In this paper we take up the idea to use artificial neural networks to model number series prediction. In contrast to [14] we apply Simple Recurrent Networks (SRNs) that are known to be a useful tool in cognitive modelling of sequence learning [4,8]. One particular advantage of SRNs in contrast to FFNs is their ability to implicit learning of the temporal characteristics of a given sequence [7].

2 Materials and Methods

In this section we present the number series used in [13], together with their results regarding the human performance on these series. Further, we describe the SRN architecture in general and the experimental setup for our modelling approach.

2.1 Number Series and Human Performance

In an experiment published in [13], 17 humans were tested on 20 different number series to evaluate the difficulty of the series. Each participant was shown each series in a random order with the request to fill in the last number. Table 1 shows the series and the result of that experiment. Additionally, we added a column giving one possible function to generate the series, where x_n denotes the n^{th} element of the series starting from $n = 1$.

2.2 Simple Recurrent Networks

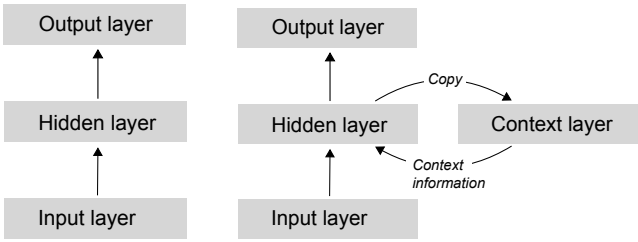
J.L. Elman was the first who described the connectionist architecture of an SRN [5]. Recurrent connections to a context layer provide the network with a dynamic memory. The processing in an SRN is *local in time* in the sense that the elements of the sequence are processed at the time of their appearance. It does not need the assumption of a fixed time window of relevant information. Further and equally important, an SRN is able to *learn an internal representation of the input sequence*.

Table 1. Results of 17 participants tested on 20 series [13]. We show the series together with the number of correct (Cor.), incorrect (Incor.) or no response (No Res.).

ID	Series	Cor.	Incor.	No Res.	Function
S01	15,12,8,11,4,7,0,3	15	0	2	$x_n = x_{n-2} - 4$
S02	148,84,52,36,28,24,22,21	12	2	3	$x_n = \frac{x_{n-1}}{2} + 10$
S03	2,12,21,29,36,42,47,51	14	1	2	$x_n = (2x_{n-1} - x_{n-2}) - 1$
S04	2,3,5,9,17,33,65,129	13	1	3	$x_n = 2x_{n-1} - 1$
S05	2,5,8,11,14,17,20,23	9	3	5	$x_n = x_{n-1} + 3$
S06	2,5,9,19,37,75,149,299	6	4	7	$x_n = 2x_{n-1} + (-1)^n$
S07	25,22,19,16,13,10,7,4	16	0	1	$x_n = x_{n-1} - 3$
S08	28,33,31,36,34,39,37,42	17	0	0	$x_n = x_{n-2} + 3$
S09	3,6,12,24,48,96,192,384	13	1	3	$x_n = 2x_{n-1}$
S10	3,7,15,31,63,127,255,511	12	3	2	$x_n = 2x_{n-1} + 1$
S11	4,11,15,26,41,67,108,175	8	1	8	$x_n = x_{n-1} + x_{n-2}$
S12	5,6,7,8,10,11,14,15	10	1	6	$x_n = x_{n-2} + \frac{n+0.5+0.5(-1)^{n-1}}{2}$
S13	54,48,42,36,30,24,18,12	16	1	0	$x_n = x_{n-1} - 6$
S14	6,8,5,7,4,6,3,5	16	0	1	$x_n = x_{n-2} - 1$
S15	6,9,18,21,42,45,90,93	14	1	2	$x_n = 2x_{n-2} + 4.5 + 1.5(-1)^{n-1}$
S16	7,10,9,12,11,14,13,16	14	0	3	$x_n = x_{n-2} + 2$
S17	8,10,14,18,26,34,50,66	13	1	3	$x_n = 2x_{n-2} - 2$
S18	8,12,10,16,12,20,14,24	17	0	0	$x_n = x_{n-2} + 3 + (-1)^n$
S19	8,12,16,20,24,28,32,36	15	0	2	$x_n = x_{n-1} + 4$
S20	9,20,6,17,3,14,0,11	16	0	1	$x_n = x_{n-2} - 3$

A standard FFN is able to develop internal representations of the input patterns in the state space of its hidden units. These internal representations are then used to produce the correct output assigned to some input (cf. Fig. 1a). The structure of a FFN and an SRN differs in one substantial point. Besides the hidden layer, a so called context layer is introduced. This layer stores the internal state of the hidden layer at the present time t. At the next time step $t+1$, this internal state is fed back to the hidden layer (cf. Fig. 1b). This simple addition has a huge effect on the processing in the network. As the context layer stores the previous internal state and provides this information to the hidden layer, the hidden units get a broader task. In an SRN the external input *and* the previous internal state have to be mapped to the desired output. The hidden layer must find a representation of some input pattern and, at the same time, find a reasonable encoding for the sequential structure of these representations.

Analytically, the processing in an SRN is defined as follows. Let $\mathbf{x}(t)$, $\mathbf{y}(t)$, and $\mathbf{z}(t)$ denote the output vectors of the input, hidden, and output layer at time t and $\mathbf{a}^x(t)$, $\mathbf{a}^y(t)$, and $\mathbf{a}^z(t)$ denote the corresponding network activation vectors. Further, $\mathbf{W}^{yx}$, $\mathbf{W}^{yy}$, and $\mathbf{W}^{yz}$ are the weight matrices for the input to hidden, context to hidden and hidden to output connections. Then the forward pass of the SRN with activation function f_{net} can be written as:



(a) FFN (b) SRN

Fig. 1. Feedforward Neural Network (FFN) (a) and Simple Recurrent Network (SRN) (b). Each box represents a network layer (set of units), and each forward arrow represents trainable connections form each unit in the lower layer to *all* units in the upper layer. The backward arrow in the SRN denotes a copy operation.

$$a_i^y(t) = \sum_j W_{ij}^{yx} x_j(t) + \sum_j W_{ij}^{yy} y_j(t-1), \tag{1}$$

$$y_i(t) = f_{\text{net}}\left(a_i^y(t)\right), \tag{2}$$

$$a_i^z(t) = \sum_j W_{ij}^{zy} y_j(t), \tag{3}$$

$$z_i(t) = f_{\text{net}}\left(a_i^z(t)\right). \tag{4}$$

2.3 Experimental Setup

One particular problem in solving number series with artificial neural networks is the scaling of the integer numbers onto an interval a network can operate on. We used the method proposed in [14]; that is, if d is the number of digits of the largest number in the series,

$$f_{\text{input}} = \frac{x}{10^d} \tag{5}$$

was used to project the integer numbers into the interval $[0, 1]$. Accordingly, the output was projected back by

$$f_{\text{output}} = x \cdot 10^d \tag{6}$$

and rounded to the nearest integer. Note that the problem at hand requires integer solutions, which poses another level of difficulty compared to usual regression task, where the real values of a function are sought.

From Table 1 one can see that the series may be solved by additions and multiplications. Therefore, we used the *linear* activation function $y = f_{\text{net}}(x) = x$ in the SRN. Instead of using random initial weights we applied an unsupervised pre-training as auto-encoder for weight initialisation. This means, the network

was trained to generate the input (numbers of the series) at the output. Such pre-training procedure can help to guide the parameters of the layers towards regions in parameter space where solutions are allowed; that is, near a solution that captures statistical structure of the input [6].

After pre-training the SRN was trained to predict the next element of the series. Therefore, the series were shown for at most 1000 training cycles, omitting the last element. After every 10 training cycles the network was tested on the complete series. If it could predict the final element of the series, it was considered to have successfully learned the rule underlying the series. For training, as for pre-training, the scaled conjugate gradient backpropagation algorithm [11] was used.

As the network should only predict the next element of the series we used one output unit. The number of input and hidden units was varied to determine the network configurations that are able to solve the task. Here the number of inputs corresponds to the "time slice" the network operates on. If it is one input unit the network has to predict the next element of the series only based on the current element, if it is two the network uses the current and the last element, and so on. At the beginning of a series, there are no previous elements. In that cases the inputs representing those elements were set to zero.

3 Results

In our experiment we trained 100 SRNs on *each* of the 20 series. Moreover, this was done for *each* combination of number of units in the input ($N_{\text{input}} = \{1, 2, \ldots, 4\}$) and hidden layer ($N_{\text{hidden}} = \{1, 2, \ldots, 5\}$). The training of 100 networks per series has two advantages. First, due to some unfavourable initial weights the gradient based learning algorithm may settle into some local minima or plateaus resulting in a bad performance of a single network. The chance to start from such weights in 100 cases is very low. Second, the training result of 100 networks tells us something about the general difficulty of the task; that is, the higher the number of networks successfully trained, the lower the task difficulty.

Table 2 shows the general result of our experiment. We give the number of series that could be learned for each network configuration. A series was considered to be learned if a *single* one of the 100 trained networks was able to predict the last element of the series.

It is remarkable that such basic architecture as an SRN consisting of only *one* input, hidden and output unit is able to reproduce 18 of the 20 number series. If we further increase the number of input units to three, all 20 series can be solved. Additionally, it becomes apparent that an increase of hidden units results in a decrease of series that could be learned: cf. Table 2, 18 if $N_{\text{input}} = N_{\text{hidden}} = 1$ against 15 if $N_{\text{input}} = 1/N_{\text{hidden}} = 5$.

To get an impression how many networks actually could learn a series, Table 3 shows the total number of networks that successfully predicted the last element of a series. As we trained 100 SRNs on each of the 20 series the maximum number

Table 2. Number of series (of 20) that could be solved by at least *one* of 100 SRNs for different combinations of input/hidden units

	hidden units				
input units	1	2	3	4	5
1	**18**	17	17	15	15
2	19	17	19	19	20
3	20	19	18	18	19
4	20	20	20	20	20

Table 3. Number of SRNs (of 2000) that could solve a series for each combination of input/hidden units

	hidden units				
input units	1	2	3	4	5
1	**1126**	989	870	837	824
2	1475	1376	1284	1203	1193
3	1467	1342	1318	1289	1237
4	**1499**	1466	1417	1388	1392

Table 4. Number of SRNs (of 100) with 4 input units that could solve the single series with different numbers of hidden units

	hidden units				
ID	1	2	3	4	5
S01	60	21	10	4	5
S02	100	100	100	100	100
S03	100	100	100	100	100
S04	98	100	100	100	100
S05	100	100	100	100	100
S06	98	92	98	99	100
S07	100	100	100	100	100
S08	2	1	4	1	2
S09	92	92	97	97	99
S10	100	100	100	100	100
S11	18	18	26	37	57
S12	100	100	100	100	100
S13	100	100	100	100	100
S14	100	62	39	19	17
S15	92	99	99	99	96
S16	4	4	5	8	5
S17	100	100	100	100	100
S18	16	69	34	22	8
S19	100	100	100	100	100
S20	19	8	5	2	3
Σ	1499	1466	1417	1388	1392

is 2000. The view on the number of successfully trained networks shows that the number of input units was the major factor in solving the series. This means, the time horizon provided to the SRN determined the chance of a successful training. The longer the history provided at the input, the higher the chance to solve a series: cf. Table 3, 1499 networks could be trained sucessfully with 4 input units while only 1126 were successful with 1 input unit. Further, Table 3 confirms the trend of Table 2. An increase of hidden units resulted in a decrease of networks that could solve a series.

From Table 1 it is clear that the different series caused different difficulties for humans. The same holds for the SRNs in our experiment. Table 4 shows the number of successfully trained networks broken down for every single series. Exemplarily we provide these numbers for the configuration with 4 input units. However, the general trend is the same for the networks with 1, 2 and 3 input units.

The comparison of Tables 1 and 4 shows that the performance of humans and networks can differ considerable on certain series. For instance series S08 could be solved by 100% of the participants in the study of [13] while only 1% – 4% of the SRNs could solve this series. For series S06 it is the other way around. Over 90% of the networks solved this series, while only 35% of the participants found the solution.

4 Discussion

Based on the SRN architecture we systematically varied the number of input and hidden units in the network to determine an adequate network configuration for the task of number series prediction. The same was done for FFN in [14]. While FFNs could solve only 18 of the 20 series, the SRNs could find a solution for all series.

In contrast to a plain feedforward processing, the hidden – context layer pair in an SRN allows the construction of a complex function (cf. Eqs. 1–4), such that SRNs consisting of only *one* input, hidden and output unit could be trained to solve 18 of 20 series. The smallest SRN configuration that was able to solve all number series consisted of only three input, one hidden and one output unit (cf. Table 2).

Even though SRNs are able to represent temporal patterns in their internal states, they benefit from a direct provision of the temporal context; that is, to show not just one, but the last two, three or four numbers of the series. This led to a significant increase of successfully trained networks with an increase of input units (cf. Table 3). The reason for the lower performance with just one input unit is the problem of vanishing error gradients when they are propagated back through time [2]. This is especially difficult in cases were the output of the network depends on inputs that occurred considerably long ago.

An increase in the number of hidden units led to a decrease of successfully trained networks (cf. Table 3). This implies that the network inputs had not to be projected into some higher dimensional space to find the function underling the series. In principle the networks can find a solution even in a very large state space; that is, high number of hidden units. However, it is more likely to get stuck in some non-optimal solution in such high dimensional space during gradient based learning.

Finally, Table 4 shows that not all series are equally well learned. This holds for humans as for SRNs. Apart from this, humans and SRNs showed different difficulties in solving a specific series. However, we cannot expect a simple architecture like an SRN to reproduce every aspect of human cognition, even if we focus on a single task like solving number series.

As we saw, a very basic SRN with three inputs, one hidden and one output unit was powerful enough to solve number series that most humans have difficulties with. Furthermore, FFNs with up to six input and up to 20 hidden units were not able to solve all of the test series [14]. From our point of view this tells us rather more about the task itself than about the processes underlying human reasoning. Nevertheless, if we compare the results of SRNs with those of FFN we can draw the conclusion that our experiment underlines the importance of recurrence in neural networks for cognitive modelling, because the recurrence is a fundamental concept in human cognition [9].

References

1. Barbounis, T., Theocharis, J., Alexiadis, M., Dokopoulos, P.: Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion 21(1), 273–284 (2006)
2. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5(2), 157–166 (1994)
3. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York (1995)
4. Cleeremans, A.: Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing. Neural Network Modeling & Connectionism. Mit Press (1993)
5. Elman, J.L.: Finding structure in time. Cognitive Science 14(2), 179–211 (1990)
6. Erhan, D., Pierre-Antoine, M., Bengio, Y., Bengio, S., Vincent, P.: The difficulty of training deep architectures and the effect of unsupervised pre-training. In: Proceedings of the 12th International Conference on Artificiall Intelligence and Statistics, pp. 153–160 (2009)
7. Glüge, S., Böck, R., Wendemuth, A.: Implicit sequence learning - a case study with a 4-2-4 encoder simple recurrent network. In: IJCCI (ICFC-ICNC), pp. 279–288 (2010)
8. Glüge, S., Hamid, O., Wendemuth, A.: A simple recurrent network for implicit learning of temporal sequences. Cognitive Computation 2, 265–271 (2010)
9. Hofstadter, D.: I am a strange loop. Basic Books, New York (2007)
10. LeFevre, J., Bisanz, J.: A cognitive analysis of number-series problems: sources of individual differences in performance. Memory and Cognition 14(4), 287–298 (1986)

11. Møller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6(4), 525–533 (1993)
12. Park, D.C.: Sunspot series prediction using adaptively trained multiscale-bilinear recurrent neural network. In: 9th IEEE International Conference on Computer Systems and Applications (AICCSA), pp. 135–139 (2011)
13. Ragni, M., Klein, A.: Predicting numbers: an AI approach to solving number series. In: Bach, J., Edelkamp, S. (eds.) KI 2011. LNCS, vol. 7006, pp. 255–259. Springer, Heidelberg (2011)
14. Ragni, M., Klein, A.: Solving number series - architectural properties of successful artificial neural networks. In: Madani, K., Kacprzyk, J., Filipe, J. (eds.) IJCCI (NCTA), pp. 224–229. SciTePress (2011)
15. Tino, P., Schittenkopf, C., Dorffner, G.: Financial volatility trading using recurrent neural networks. IEEE Transactions on Neural Networks 12(4), 865–874 (2001)

Detection Method for Phase Synchronization in a Population of Spiking Neurons

Manuel Lopez and Francisco B. Rodríguez

GNB. E.T.S. Ingeniería Informática, Universidad Autonóma de Madrid,
Cra. Colmenar Viejo, Km. 15, 28049 Madrid, Spain
mlopezm@acm.org, f.rodriguez@uam.es

Abstract. Currently there are many methods to detect synchronization, each of them trying to extract some specific aspects or oriented to specific number or type of signals. In this paper, we present a new method to detect synchronization for multivariate signals, computationally light and not requiring a combinatorial number of operations on signals differences.The method is based on the Hilbert transform of the signals, which provides their instantaneous phases. The distribution of phases for all signals at a specific time is assimilated to a probability distribution. In this way, we obtain a sequence of probability distributions (one per time unit). Computing the entropy of the probability distributions we get finally a function of entropies along time. The average value of this final function provides a good estimate of the synchronization level of the multivariate signals ensemble, and the function itself can be used as a signature (descriptive function) of the whole multidimensional ensemble dynamics.

Keywords: synchronization detection, signal phase, multivariate signals.

1 Introduction

The study of synchronization is a main topic of current research in different fields: Computational neuroscience [1], Temporal synchronization of image sequence [10], Data mining in multivariate data, Information retrieval by content [9], Analysis of geologic and atmospheric data [4], Human movements, gaits and gesture recognition [5], Financial applications (trends, similarities,..) [8,2], Genetics (DNA,..) [3], etc ...

Synchronization is defined by Webster dictionary as: (1) happening, existing, or arising at precisely the same time, (2) recurring or operating at exactly the same periods.This apparently easy definition hides some troubles in its practical usage due to: (1) inherent signal noise, (2) imperfect signal coincidences, (3) alternatives in signal parameter used to establish signals similarity (instantaneous value, phase, frequency,), (4) establishing similarity between more than two signals, (5) similarity between two signals ensembles, (6) patterns recurrence (which pattern? spatio-temporal?), These difficulties imply more efforts dedicated to provide new and better methods for synchronization detection mechanisms.

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part I, LNCS 7930, pp. 421–431, 2013.

These efforts are justified due to the importance of synchronization in the above mentioned fields. Particularly, in neuroscience, synchronization is been studied as one of the main elements that the brain uses in the integration of dissimilar information to form a coherent perception (binding problem) [6] and as an important mechanism for information coding and learning [7].

Considering all these aspects we propose in this paper a new method to detect synchronization for multivariate signals. The method is computationally low demanding as it is not based on a combinatorial number of differences between signals, but, in a time-dependent entropy function calculated on the signals probability phase distributions. Therefore, the number of main-step computations is made linear with the total number of time units of the signals: $O(T)$, where T is the total time length of the signals .

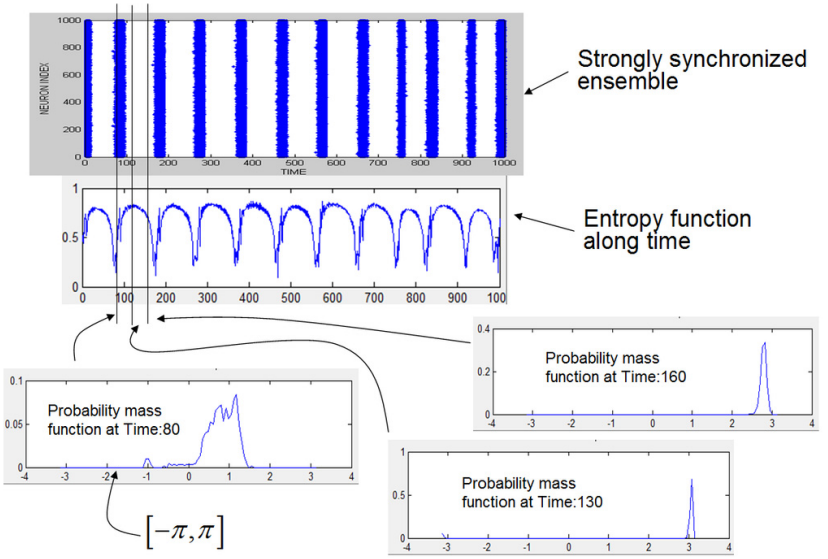


Fig. 1. Relationship between signal ensemble, entropy sequence function $\gamma_S(t)$ and probability mass functions p_t for different times

2 Multivariate Synchronization Detection Method

In this section is presented the detection method. First it is presented the Hilbert transform which allows getting the instantaneous signal phase, following with a detailed description of the proposed synchronization detection method.

2.1 Hilbert Transform

Assuming a discrete-time, continuous-valued function: $x = \{x_1, x_2, ..x_T\}$; $x_i \in \mathbf{R}$ the analytic function of $x(t)$ is defined as [11,3]:

$$\xi(t) = x(t) + x_H(t) = A(t)exp[i\phi(t)], \tag{1}$$

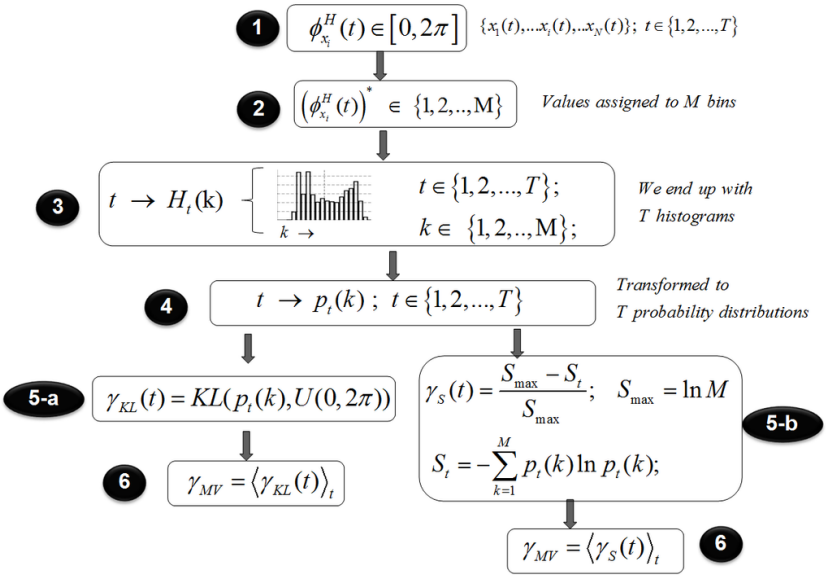


Fig. 2. Synopsis of method computation

$$\phi(t) = \tan^{-1}\left[\frac{x_H(t)}{x(t)}\right],\ x_H(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}\, d\tau, \tag{2}$$

where $x(t)$ is the signal and $x_H(t)$ is the Hilbert transform of it, and P.V. indicates that the integral is taken in the sense of the Cauchy principal value. In Eq. 1, $\xi(t)$ is the analytic function and the functions $A(t)$ and $\phi(t)$ define the instantaneous phase and instantaneous amplitude of function $x(t)$.

2.2 Method Description

To start with we assume having a set of N interconnected nodes (neurons). The activity of the N nodes produce N signals which we consider to be sampled at regular time intervals producing T samples per signal. Ending up with an activity matrix $\mathbf{X}$ of dimension $N \times T$ with the sampled values of all signals:

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^T \\ \dots & . & . & \dots \\ x_N^1 & x_N^2 & \dots & x_N^T \end{bmatrix} = \left(\boldsymbol{x}^1, \boldsymbol{x}^2, \dots, \boldsymbol{x}^T\right) \tag{3}$$

Activity matrix $\mathbf{X}$ can be also represented as an array of column vectors $\boldsymbol{x}^i$, each column vector having the activity of all nodes at time i .

Computing the Hilbert transform of all signals in matrix $\mathbf{X}$ will provide matrix $\mathbf{X_H}$ (Hilbert transform along rows of $\mathbf{X}$), from which we can compute the instantaneous matrix of signals phases $\mathbf{\Phi}$:

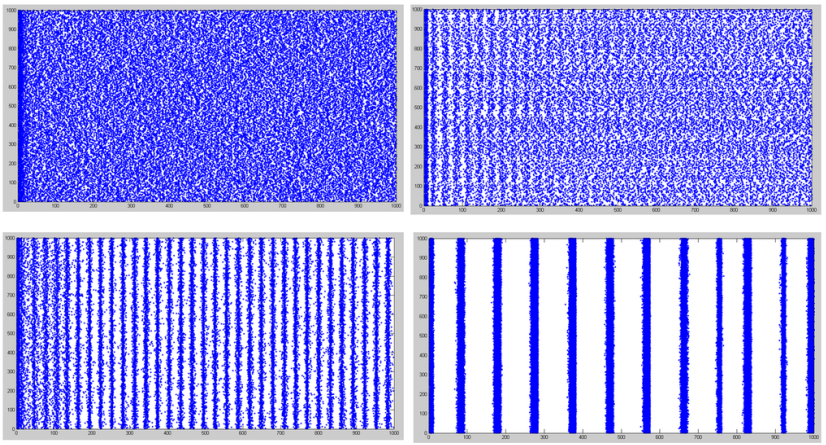


Fig. 3. Regular synchronous signal. We show four ensembles used for RS synchronization, from left to right and up to bottom: random (no synchronization), weakly, moderately and strongly synchronized ensembles.

$$\mathbf{X_H} = \begin{bmatrix} x_{1H}^1 & x_{1H}^2 & \dots & x_{1H}^T \\ \dots & . & . & \dots \\ x_{NH}^1 & x_{NH}^2 & \dots & x_{NH}^T \end{bmatrix} \tag{4}$$

$$\mathbf{\Phi} = \begin{bmatrix} \tan^{-1}\left(\frac{x_{1H}^1}{x_1^1}\right) & \tan^{-1}\left(\frac{x_{1H}^2}{x_1^2}\right) & \dots & \tan^{-1}\left(\frac{x_{1H}^T}{x_1^T}\right) \\ \dots & . & . & \dots \\ \tan^{-1}\left(\frac{x_{NH}^1}{x_N^1}\right) & \tan^{-1}\left(\frac{x_{NH}^2}{x_N^2}\right) & \dots & \tan^{-1}\left(\frac{x_{NH}^T}{x_N^T}\right) \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} \phi_1^1 & \phi_1^2 & \dots & \phi_1^T \\ \dots & . & . & \dots \\ \phi_N^1 & \phi_N^2 & \dots & \phi_N^T \end{bmatrix} = \left(\boldsymbol{\phi}^1, \boldsymbol{\phi}^2, \dots, \boldsymbol{\phi}^T\right) \tag{6}$$

Again, matrix $\mathbf{\Phi}$ can be represented as an array of column vectors $\boldsymbol{\phi}^i$, each column vector having the instantaneous phases of all nodes at time i .

Considering the phase values for each column vector $\boldsymbol{\phi}^i$: we build its histogram, assigning the range of phase values $[0, 2\pi]$ into M bins.

Finally we will have T histograms (one per column vector $\boldsymbol{\phi}^i$) where in bin k of histogram i one will have the total number of times that any signal phase at time i has a value between $2\pi\frac{(k-1)}{M}$ and $2\pi\frac{k}{M}$ radians.

Normalizing the histogram to have a total area of 1, we end up with T probability mass functions: p_t , one per time increment:

$$t \to p_t(k),\ t = \{1, 2, \dots, T\},\ k = \{1, 2, \dots, M\}. \tag{7}$$

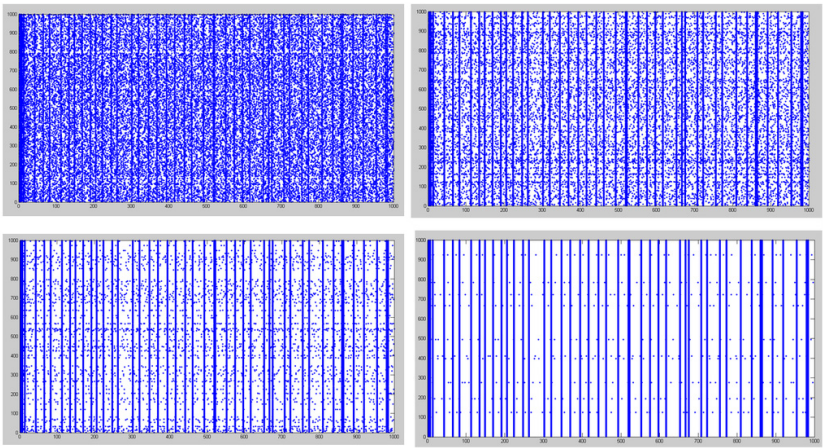


Fig. 4. Irregular synchronous signals with 30%, 70%, 90%, and 99% as percentage of identical signals in the ensemble (from left to right and up to bottom)

If the distribution of the probability mass functions is similar to a uniform distribution implies that the signals are not synchronized and the more the distribution is departed from the uniform distribution implies that the signals are more synchronized, being the limit a delta Dirac distribution with a single phase value for all signals.

We define a synchronization index, based in applying a measure to each probability distribution: $p_t(k)$, to finally have a single number that represents how close is the distribution to a uniform distribution. In this paper, we have used the Entropy.

Applying the Entropy operator to each probability mass function p_t, we transform a sequence of probability distributions $\mathbf{P} = [p_1, p_2, \ldots, p_T]$ into a sequence of scalars $\mathbf{S} = [S_1, S_2, \ldots, S_T]$, where:

$$S_t = -\sum_{k=1}^{M} p_t(k) \ln p_t(k), \;\; t = \{1, 2, \ldots, T\}. \tag{8}$$

Normalizing the entropies, with respect to the maximum entropy which is given by the uniform distribution ($\ln M$), we transform the entropies S_t into normalized entropies $\gamma_S(t)$ which range is $[0, 1]$:

$$\gamma_S(t) = \frac{S_{max} - S_t}{S_{max}}; \; S_{max} = \ln M \tag{9}$$

The final result is that the ensemble activity matrix $\mathbf{X}$ in Eq. 1 has been transformed into a sequence of scalars $[\gamma_S(1), \gamma_S(2), \ldots, \gamma_S(T)]$ of length T, which we call the **entropy sequence function**. This sequence of scalars actually represents a function of time of the normalized entropies associated to per-time-sliced phase values distributions along all nodes in the ensemble.

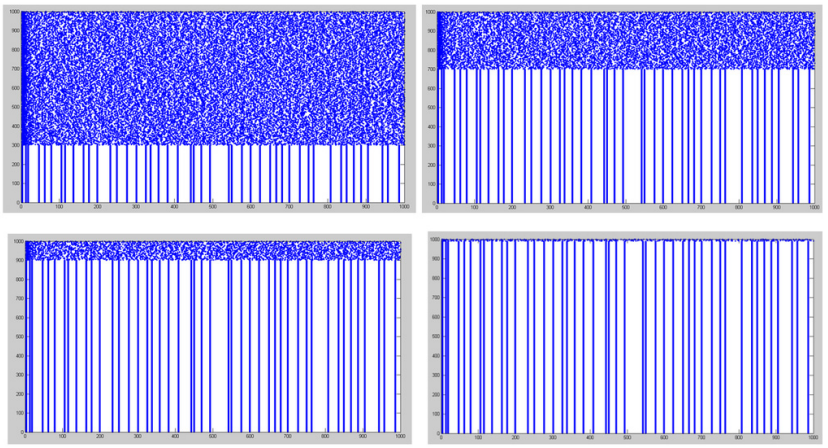


Fig. 5. Irregular synchronous signals forming clusters of contiguous identical signals. We show from left to right and up to bottom four ensembles with 30%, 70%, 90%, and 99% as percentage of identical signals.

From the entropy sequence function we can provide a summary value (a single scalar) which can represent the complete activity of the ensemble. Several *aggregation* functions can be taken, but we have used the simple average function along time:

$$\gamma_{MV} = \langle \gamma_S(t) \rangle_t \tag{10}$$

This final value is the **multivariate synchronization index** presented in this paper.

In Fig 1 is presented the relationship between the signal ensemble, the entropy sequence function: $\gamma_S(t)$ and probability mass functions: p_t . We can see that the points in time of maximum entropy corresponds with quasi Dirac deltas in the probability distributions.

In Fig 2 is presented a synopsis of the computations performed: In step 1, the Hilbert transform of each signal is calculated. In step 2 we discretize the phase values to M possible values (M bins). In step 3 we build a histogram per time increment, the histogram gives the relative frequencies of each possible value: $\{1, 2, \ldots, M\}$, along all the nodes. Normalizing the histograms we construct the corresponding probability mass functions in step 4. From there, we can use the Entropy to translate probability distributions to scalar values, which is done in steps 5. Note that we could use any alternative operator in order to translate probability distributions to scalar values, like for instance the Kullback-Leibler(KL) [12] divergence. In step 6 we provide an aggregate summary of the Entropies or KL divergence, which are functions of time; to do so we use the average along time.

An important secondary result is the possible use of the function $\gamma_S(t)$ as a signature (single representative function) of the whole multidimensional

	IS Signals
Random	0.28745
IS-30%	0.39439
IS-70%	0.67452
IS-90%	0.87018
IS-99%	0.98441
IS-100%	1

	IS Signals in clusters
Random	0.28745
IS-30%	0.39423
IS-70%	0.67411
IS-90%	0.87006
IS-99%	0.98431
IS-100%	1

	RS Signals
Random	0.28745
RS-Weakly Synchronized	0.42574
RS-Moderately Synchronized	0.46936
RS-Strongly Synchronized	0.68369

Fig. 6. Multivariate synchronization index (γ_{MV}) results for different types of signals: RS, IS and IS-C (Section 3). At the bottom of each data set is presented a function representation of values (from top to bottom).

ensemble dynamics. So, these functions can be used to compare the similarity of two complete ensembles of signals.

In the following sections we will provide evidence of the suitability of the method to detect synchronization.

3 Data Test Signals

To test the method we have used three types of signals with increasing level of synchrony in each group.

3.1 Regular Synchronous (RS)

These signals have been produced with a specifically developed Matlab toolbox starting with a non-connected set of 1000 neurons with a random input signal for each of the neurons to produce a random signals ensemble. Then it has been incrementally increased the coupling weight and degree of connectivity of the network, producing three additional signals ensembles with: weak, medium and strong synchronization.

To create the simulation we have used the Izhikevich neuron model [13,14]. This model is a phenomenological spiking model as computationally efficient as the integrate-and-fire model and that depending only on four parameters can reproduce the spiking and bursting behavior of known types of cortical neurons. For this specific simulation we have created 1000 excitatory integrator neurons with a random connectivity network.

In Fig 3 are presented the four ensembles used for RS synchronization, from left to right and up to bottom: random (no synchronization), weak, medium and strong synchronization ensembles.

In all cases the signals present a periodic or quasi-periodic recurrence activity (except the first which is random). The signal coming from each neuron is different to all other signals but the global activity forms a periodic dynamic.

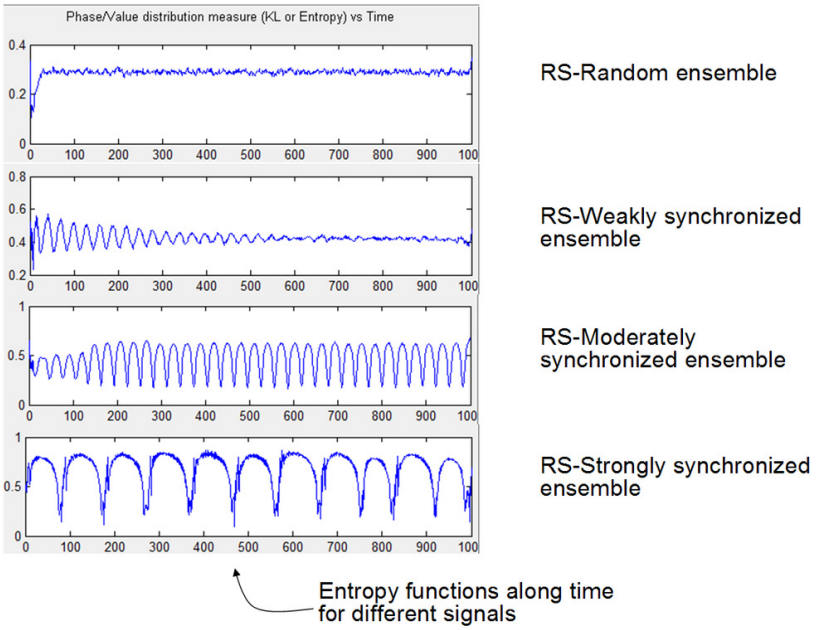


Fig. 7. Entropy sequence functions for the four RS signals. The x-axis is time.

3.2 Irregular Synchronous (IS)

These signals have been produced starting from the same random ensemble of 1000 neurons used in the RS case. From this ensemble, we have randomly selected a neuron signal and replaced a percentage of other neurons signals with the chosen one. In this way, it has been produced several ensembles with a different percentage of repetition of the one single signal chosen for substitution. There are ensembles with 30%, 70%, 90%, 99% and 100% as percentage of identical signals in the ensemble. For example, the 100% repetition case is an ensemble with all neurons signals been identical.

The repetition of equal signals in the ensemble is done in a random manner, meaning that the replacing signal occupies random positions in the signals matrix rows.

In Fig 4 are presented several ensembles (from left to right and up to bottom) for 30%, 70%, 90% and 99% repetition percentage. The figure presents the spikes activity of the signals, being time in horizontal and nodes in vertical.

These signals have been produced using a specifically developed Matlab code; they have not been generated by the Matlab toolbox as part of the simulation of a dynamical system, as for the RS signals.

3.3 Irregular Synchronous in Cluster (IS-C)

Similar to IS signals but with equal signals placed in contiguous positions. In Fig 5 are presented several ensembles (from left to right and up to bottom) for 30%, 70%, 90% and 99% repetition percentage of the IS signals in cluster.

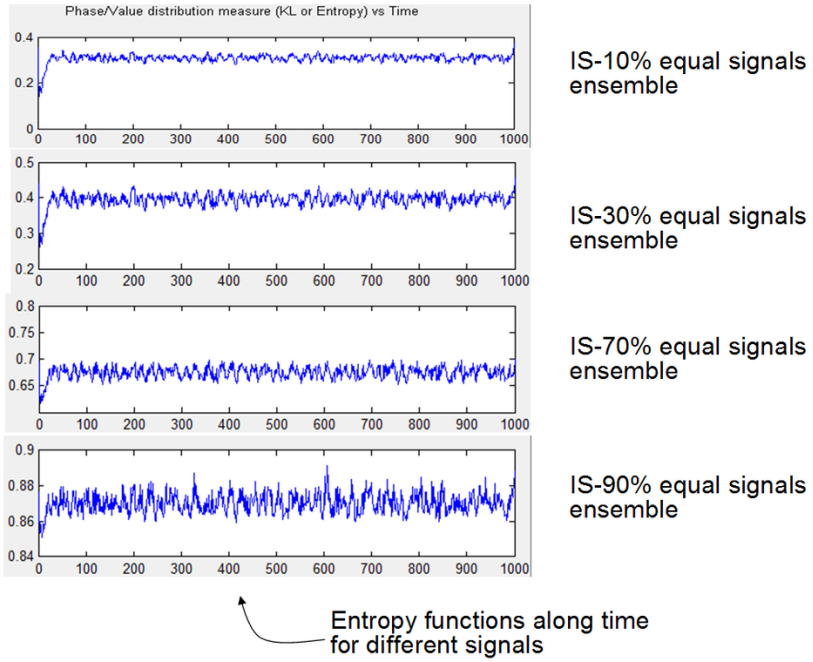


Fig. 8. Entropy sequence functions for the four IS signals. The x-axis is time.

4 Experimental Results

In this section we present the numerical results obtained after applying the multivariate index of section 2.2 to the signals in section 3. The results are presented in Fig 6. In this figure we can observe that the evolution of the phase index is monotonically increasing as synchronization increases in the ensembles, providing a convenient measure of synchronization between ensembles, even of different types (IS vs. RS). In Fig 7 are shown the entropy sequence functions for the four RS signals in section 3. The four diagrams correspond to (from up to bottom): random (no synchronization), weakly, moderately and strongly

synchronized signals. In Fig 8 is presented a similar diagram but for the four IS signals of sections 3. We can see, observing these figures, how the differences between the IS and RS signals are expressed in the shape of the entropy functions ($\gamma_S(t)$). The IS signals show an almost constant entropy sequence function only modulated by noise, with an average value increasing as the signals ensemble increase in synchrony. Meanwhile, the RS signals present and oscillating shape, centered around an almost constant value but with the oscillation amplitudes increasing as the synchronization increases. Considering the above mentioned properties, the entropy function can be used as a one-dimensional signature function to represent the whole signals ensemble.

5 Conclusions

It has been presented a synchronization detection method for multivariate signals which is suitable to detect the synchronization level of a signals ensemble. The method grows linearly in computational demands with the increase in number of signals and sampling time steps. It is not based in a combinatorial growing number of signals differences.

The method is presented in detail and the results show its adequacy to detect synchronization for different types of signals.

The method gives two main results:

- **A scalar value (phase index)**: This provides a single aggregate of synchronization level.
- **A time function of entropies**: This provides a signature of the ensemble behavior good to capture the similarity signals level along time. This function can be used to compare two ensembles of signals.

Acknowledgements. This work was supported by the Spanish Government projects TIN2010-19607.

References

1. Dauwels, J., Vialatte, F., Cichocki, A.: A Comparative Study of Synchrony Measures for the Early Detection of Alzheimers Disease Based on EEG. Elsevier, NeuroImage 49, 668–693 (2010)
2. Granger, C.W.J.: Testing for causality: A personal viewpoint. Journal of Economic Dynamics and Control 2, 329–352 (1980)
3. Sellers, P.H.: On the theory and computation of evolutionary distances. SIAM J. Appl. Math. 26, 787–793 (1974)
4. Plaut, G., Vautard, R.: Spells of Low-Frequency Oscillations and Weather Regimes in the Northern Hemisphere. Journal of Atmospheric Sciences 51(2), 210–236 (1993)
5. Gillian, N., Knapp, R.B., O'Modhrain, S.: Recognition of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping. In: Proceedings of NIME 2011, Oslo, Norway (May 2011)

6. Dong, Y., Mihalas, S., Qiu, F., von der Heydt, R., Niebur, E.: Synchrony and the binding problem in macaque visual cortex. Journal of Vision 8(7), 1–16 (2008)
7. Borisyuk, R., Borisyuk, G.: Information coding on the basis of synchronization of neural activity. BioSystems 40, 3–10 (1997)
8. Liu, X.F., Tse, C.K.: A complex network perspective of world stock markets: synchronization and volatility. International Journal of Bifurcation and Chaos 22(6) (2012)
9. Müller, M.: New Developments in Music Information Retrieval. In: Proceedings of the 42nd AES Conference (2011)
10. Dexter, E., Perez, P., Laptev, I., Junejo, I.N.: Multi-view Synchronization of Human Actions and Dynamic Scenes. In: VISAPP 2009: Proceedings 4th International Conference on Computer Vision Theory and Applications, vol. 2, pp. 383–391 (2009)
11. Pereda, E., Quiroga, R.Q., Bhattacharya, J.: Nonlinear multivariate analysis of neurophysiological signals. Progress in Neurobiology 77, 1–37 (2005)
12. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
13. Izhikevich, E.M.: Simple Model of Spiking Neurons. IEEE Trans. Neural Networks 14(6), 1569–1572 (2003)
14. Izhikevic, E.M.: Which Model to Use for Cortical Spiking Neurons? IEEE Transactions on Neural Networks 15, 1063–1070 (2004)

Author Index

Achab, Anissa I-396
Adell, Jordi I-121
Agullo, C. I-342
Aliaga, R.J. II-413
Andina Diego I-1
Álvarez, Ignacio II-258
Álvarez-Meza, A.M. II-365
Álvarez-Sánchez, José Ramón I-37, II-45
Alvarez-Santos, Víctor II-92, II-112
Analide, C. II-200
Anaya, Antonio R. I-235
Andina, Diego I-9, I-181
Ansoategui, Igor II-441
Arias-Londoño, J.D. I-201, I-274, I-283
Arini, Pedro I-350
Arnau Prieto, Raúl II-32, II-45
Avila, C. II-140
Ayerdi, Borja II-122
Azorín, José María I-131, I-151

Bäck, Thomas II-1
Barakova, Emilia I-161
Barbi, M. I-48
Barone, Valeria I-350
Barrientos C., Fernanda I-245
Barros-Loscertales, A. II-140
Bartolomé, Elena I-74
Bastida-Jumilla, María-Consuelo I-386, II-241, II-250
Belalcazar-Bolaños, E.A. I-283
Bellas, Francisco II-56, II-102
Benatchba, Karima I-396
Benchaib, Y. I-181
Berenguer-Vidal, Rafael II-230, II-250
Berlanga, Antonio II-149
Bermudez, Maria II-404
Bologna, Guido I-83
Bonomini, María Paula I-350
Boutouchent, Amira I-396

Caamaño, Pilar II-56
Cabido, R. II-286, II-296
Campoy-Muñoz, Pilar II-326
Canedo-Rodriguez, A. II-92, II-112
Capriles, Jose Rafael I-121
Carmona, Enrique J. II-11
Caro, Stéphane II-451
Castellanos-Domínguez, C.G. I-254, I-264, I-274, I-283, II-365
Castello Paiva, Daniel I-1
Castillo, José Carlos II-220
Castro Hoyos, C. I-254
Cerquera, Alexander I-293
Chae, Hyunuk I-103
Chillemi, S. I-48
Chyzhyk, Darya II-131, II-140
Cilla, Rodrigo II-149
Collazos, Antonio II-169
Colodro-Conde, Carlos II-375, II-385
Contreras, Ricardo I-313, I-323
Crawford, Broderick II-451
Cuadra-Troncoso, José Manuel II-32, I-37, II-45, II-190
Cuenca-Asensi, Sergio I-332

Dahamni, Fodil I-396
de la Paz López, Félix I-37, II-32
Delgado, María Luisa II-169, II-286
de Lope, Javier II-66, II-73
de Pablos Álvaro, Marta I-9
de Santos-Sierra, D. I-342
Díaz-García, José Javier II-385
Díaz-López, Estela I-191
Di Garbo, A. I-48
Duro, Richard J. II-56, II-102

Emmerich, Michael II-1
Etxeberria-Agiriano, Ismael II-441

Faíña, Andrés II-56, II-102
Fdez-Vidal, Xosé R. II-345
Fe, Jorge II-413
Fernández, Eduardo I-37, I-131, I-151, I-332, I-342
Fernández, Miguel Angel II-210
Fernández-Caballero, Antonio II-159, II-179, II-210, II-220, II-306

Fernandez-Delgado, M. II-92
Fernandez-Gauna, Borja II-441
Fernández-López, David II-169, II-286, II-296
Fernandez-Luque, Francisco II-355
Ferrández-Vicente, José Manuel I-37, I-212, I-332, II-375, II-385
Ferreira C., Anita I-245
Ferreiroa, Rubén I-26
Fuentes, José Alberto II-210
Fuentes Brea, Juan Pablo II-66

Gadea, R. II-413
Galindo, Cipriano I-141
Gamallo, C. II-92
Garcerán-Hernández, Vicente I-303, I-358
García-Díaz, Antón II-345
Garcia-Perez, Joaquin D. I-171
García-Rodríguez, Coral II-190
García-Sáiz, Tomás I-191
García-Vega, S. I-264
Garrigós-Guerrero, Fco. Javier I-37, II-375, II-385
Glüge, Stefan I-412
Gomez, Juan Diego I-83
Gómez Ávila, Gabriel Fatíh II-81
Gómez-Vilda, Pedro I-74, I-212
González, Pascual II-159
Gonzalez-Jimenez, Javier I-141
González-López, Antonio I-386, II-241
González Pacheco Oceguera, José Eduardo II-81
González-Rodríguez, Inés I-376
Gonzalo-Fonrodona, Isabel I-368
Górriz, Juan Manuel I-406, II-258, II-266, II-278
Graña, Manuel II-122, II-131, II-140, II-317, II-336, II-441
Guerrero-Gómez-Olmedo, Ricardo II-306
Gutiérrez, P.A. II-326
Gutierrez-Castaneda, Carlos I-141

Hamdad, Leila I-396
Hasegawa, Tsutomu I-103, I-113
Hervás-Martínez, C. II-326
Hortal, Enrique I-131, I-151
Huerta, Ramon I-16

Iáñez, Eduardo I-131, I-151
Iglesias, Roberto II-92, II-112
Illán, I.A. I-406, II-266
Ingallina, Fernando I-350

Kok, Joost II-1
Kurazume, Ryo I-103, I-113
Kuwahata, Shunya I-103

Lafuente-Arroyo, Sergio II-395
Langseth, Helge I-225
Larrey-Ruiz, Jorge I-386
Leboeuf-Pasquier, Jérôme II-81
Leborán Alvarez, Víctor II-345
Legaz-Aparicio, Álvar II-230
Londral, Ana Rita M. I-212
Lopez, Manuel I-421
López, María T. II-179, II-220
Lopez-Guede, Jose Manuel II-317
López-Sastre, Roberto J. II-179, II-306, II-395
Lorente, V. I-37
Lourens, Tino I-161
Luque, Manuel I-235

Maldonado-Bascón, Saturnino II-306, II-395
Maravall, Darío II-66, II-73
Marcano-Cedeño, Alexis I-181
Marchionni, Luca I-121
Marin-Alonso, Oscar I-171
Marques, Ion II-336
Marquez, David I-225
Martínez, Jonatan II-159
Martínez-Álvarez, Antonio I-332
Martínez-Álvarez, José Javier II-375, II-385
Martinez Mozos, Oscar I-103
Martínez-Murcia, Francisco Jesús I-406, II-266
Martínez-Rams, Ernesto A. I-303, I-358
Martínez-Tomás, Rafael I-191, II-190, II-395
Menchón-Lara, Rosa-María II-230, II-241, II-250
Miguel-Tomé, Sergio II-461
Mingo, Jack Mario II-73
Molina, José Pascual II-159
Molina, José M. II-149
Molina-Casado, José María II-11

Molina-Ruíz, R.M. I-191
Monfroy, Eric II-451
Monteagudo, Ángel II-21
Montemayor, Antonio S. II-169, II-286, II-296
Montero, Aaron I-16
Montero, Francisco II-210
Morales-Sánchez, Juan I-386, II-250
Moreno, A. Belén II-431
Morooka, Ken'ichi I-103, I-113
Muñoz, Miguel A. II-159
Murillo-Rendón, S. I-254, I-274

Navarro Santosjuanes, Israel II-32, II-45
Neil, Martin I-225
Nöth, Elmar I-201, I-283

Olivares, Teresa II-200
Oliver, Miguel II-210
Olmedo-Payá, Andrés I-332, I-342
Oportus, Eduardo I-313
Orjales, Felix II-102
Orozco-Arroyave, J.R. I-201, I-283
Ortega, Jaime I-323
Ortiz García, Andrés II-278
Oterino, Fernando II-317

Padilla, Pablo II-258
Pages, Jordi I-121
Palacios, Juan José I-376
Palacios-Alonso, Daniel I-74
Pantrigo, Juan José II-169, II-286, II-296
Pardo, Xosé M. II-112, II-345
Patricio, Miguel A. II-149
Peluffo-Ordóñez, D. I-264, I-274
Pinninghoff, M. Angélica I-313, I-323
Porras, Miguel A. I-368
Puente, Jorge I-376
Puertas, Jeronimo II-404
Pun, Thierry I-83
Puntonet, C.G. I-406, II-266
Pyo, YoonSeok I-103

Rabuñal, Juan R. II-404
Ramírez, Javier I-406, II-258, II-266, II-278
Ramos-Hernanz, Jose Antonio II-317
Rebolo, Rafael II-385
Regueiro, C.V. II-92, II-112
Rincón-Zamorano, Mariano I-191
Rodellar-Biarge, Victoria I-74, I-212
Rodriguez, Alvaro II-404
Rodríguez, Francisco B. I-16, I-421
Ropero Peláez, Francisco Javier I-1
Royo, F. II-200
Ruiz, Ramón II-355
Ruiz-Fernández, Daniel I-171

Salas-González, Diego I-406, II-258, II-278
Salcedo L., Pedro I-245, I-313
Sánchez, Ángel II-431
Sánchez, Eduardo I-26
Sánchez-Oro, J. II-296
Sancho-Gómez, José Luis II-241
Santos Reyes, José II-21
Santos-Saavedra, D. II-92
Savio, Alexandre II-122
Segovia, F. I-406
Serrano-Cuerda, Juan II-179, II-220
Silva, F. II-200
Sokolova, Marina V. II-220
Sonoki, Shuji I-113
Soriano, Antonio I-171
Soto, Ricardo II-451
Soto-Sanchez, C. I-342

Tapus, Adriana I-94
Termenon, M. II-140
Thilakarathne, Dilhan J. I-62
Timón-Reina, Santiago I-191
Toledo-Moreo, Fco. Javier II-375, II-385
Toledo-Moreo, Rafael II-385
Tomé, Hilario I-121
Torres-Alegre, Santiago I-181
Treur, Jan I-62
Triana, Nayid I-293
Tribaldos, Pablo II-179
Tsuji, Tokuo I-103
Tubío-Araujo, Óscar Manuel II-385

Úbeda, Andrés I-131, I-151

Vallone, F. I-48
van Veen, Barry II-1
Vargas-Bonilla, J.F. I-201, I-283
Vela, Camino R. I-376
Velásquez-Martínez, Luisa F. II-365

Vélez, JoséF. II-431
Verdú-Monedero, Rafael II-230
Vergara, M.A. II-200
Vieru, Alexandra-Maria I-94
Vilar, Rafael II-421
Villó-Pérez, Isidro II-375, II-385
Wendemuth, Andreas I-412
Yáñez, Andrea I-313
Yang, Zhiwei II-1
Zapata, Juan II-355, II-421