

Randall Shumaker (Ed.)

LNCS 8021

# Virtual, Augmented and Mixed Reality

Designing and Developing Augmented  
and Virtual Environments

5th International Conference, VAMR 2013  
Held as Part of HCI International 2013  
Las Vegas, NV, USA, July 2013, Proceedings, Part I

1  
Part I



 Springer

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Alfred Kobsa

*University of California, Irvine, CA, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*TU Dortmund University, Germany*

Madhu Sudan

*Microsoft Research, Cambridge, MA, USA*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Gerhard Weikum

*Max Planck Institute for Informatics, Saarbruecken, Germany*

Randall Shumaker (Ed.)

# Virtual, Augmented and Mixed Reality

Designing and Developing Augmented  
and Virtual Environments

5th International Conference, VAMR 2013  
Held as Part of HCI International 2013  
Las Vegas, NV, USA, July 21-26, 2013  
Proceedings, Part I



Springer

## Volume Editor

Randall Shumaker  
University of Central Florida  
Institute for Simulation and Training  
3100 Technology Parkway, Orlando, FL 32826, USA  
E-mail: shumaker@ist.ucf.edu

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-642-39404-1 e-ISBN 978-3-642-39405-8  
DOI 10.1007/978-3-642-39405-8  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013942173

CR Subject Classification (1998): H.5, H.3, H.1, I.3.7, H.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

*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)



# Foreword

The 15th International Conference on Human–Computer Interaction, HCI International 2013, was held in Las Vegas, Nevada, USA, 21–26 July 2013, incorporating 12 conferences / thematic areas:

Thematic areas:

- Human–Computer Interaction
- Human Interface and the Management of Information

Affiliated conferences:

- 10th International Conference on Engineering Psychology and Cognitive Ergonomics
- 7th International Conference on Universal Access in Human–Computer Interaction
- 5th International Conference on Virtual, Augmented and Mixed Reality
- 5th International Conference on Cross-Cultural Design
- 5th International Conference on Online Communities and Social Computing
- 7th International Conference on Augmented Cognition
- 4th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management
- 2nd International Conference on Design, User Experience and Usability
- 1st International Conference on Distributed, Ambient and Pervasive Interactions
- 1st International Conference on Human Aspects of Information Security, Privacy and Trust

A total of 5210 individuals from academia, research institutes, industry and governmental agencies from 70 countries submitted contributions, and 1666 papers and 303 posters were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of Human–Computer Interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Randall Shumaker, contains papers focusing on the thematic area of Virtual, Augmented and Mixed Reality, and addressing the following major topics:

- Developing Augmented and Virtual Environments
- Interaction in Augmented and Virtual Environments
- Human-Robot Interaction in Virtual Environments
- Presence and Tele-presence

The remaining volumes of the HCI International 2013 proceedings are:

- Volume 1, LNCS 8004, Human–Computer Interaction: Human-Centred Design Approaches, Methods, Tools and Environments (Part I), edited by Masaaki Kurosu
- Volume 2, LNCS 8005, Human–Computer Interaction: Applications and Services (Part II), edited by Masaaki Kurosu
- Volume 3, LNCS 8006, Human–Computer Interaction: Users and Contexts of Use (Part III), edited by Masaaki Kurosu
- Volume 4, LNCS 8007, Human–Computer Interaction: Interaction Modalities and Techniques (Part IV), edited by Masaaki Kurosu
- Volume 5, LNCS 8008, Human–Computer Interaction: Towards Intelligent and Implicit Interaction (Part V), edited by Masaaki Kurosu
- Volume 6, LNCS 8009, Universal Access in Human–Computer Interaction: Design Methods, Tools and Interaction Techniques for eInclusion (Part I), edited by Constantine Stephanidis and Margherita Antona
- Volume 7, LNCS 8010, Universal Access in Human–Computer Interaction: User and Context Diversity (Part II), edited by Constantine Stephanidis and Margherita Antona
- Volume 8, LNCS 8011, Universal Access in Human–Computer Interaction: Applications and Services for Quality of Life (Part III), edited by Constantine Stephanidis and Margherita Antona
- Volume 9, LNCS 8012, Design, User Experience, and Usability: Design Philosophy, Methods and Tools (Part I), edited by Aaron Marcus
- Volume 10, LNCS 8013, Design, User Experience, and Usability: Health, Learning, Playing, Cultural, and Cross-Cultural User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 8014, Design, User Experience, and Usability: User Experience in Novel Technological Environments (Part III), edited by Aaron Marcus
- Volume 12, LNCS 8015, Design, User Experience, and Usability: Web, Mobile and Product Design (Part IV), edited by Aaron Marcus
- Volume 13, LNCS 8016, Human Interface and the Management of Information: Information and Interaction Design (Part I), edited by Sakae Yamamoto
- Volume 14, LNCS 8017, Human Interface and the Management of Information: Information and Interaction for Health, Safety, Mobility and Complex Environments (Part II), edited by Sakae Yamamoto
- Volume 15, LNCS 8018, Human Interface and the Management of Information: Information and Interaction for Learning, Culture, Collaboration and Business (Part III), edited by Sakae Yamamoto
- Volume 16, LNAI 8019, Engineering Psychology and Cognitive Ergonomics: Understanding Human Cognition (Part I), edited by Don Harris
- Volume 17, LNAI 8020, Engineering Psychology and Cognitive Ergonomics: Applications and Services (Part II), edited by Don Harris
- Volume 19, LNCS 8022, Virtual, Augmented and Mixed Reality: Systems and Applications (Part II), edited by Randall Shumaker

- Volume 20, LNCS 8023, Cross-Cultural Design: Methods, Practice and Case Studies (Part I), edited by P.L. Patrick Rau
- Volume 21, LNCS 8024, Cross-Cultural Design: Cultural Differences in Everyday Life (Part II), edited by P.L. Patrick Rau
- Volume 22, LNCS 8025, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Healthcare and Safety of the Environment and Transport (Part I), edited by Vincent G. Duffy
- Volume 23, LNCS 8026, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Human Body Modeling and Ergonomics (Part II), edited by Vincent G. Duffy
- Volume 24, LNAI 8027, Foundations of Augmented Cognition, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
- Volume 25, LNCS 8028, Distributed, Ambient and Pervasive Interactions, edited by Norbert Streitz and Constantine Stephanidis
- Volume 26, LNCS 8029, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 27, LNCS 8030, Human Aspects of Information Security, Privacy and Trust, edited by Louis Marinos and Ioannis Askoxylakis
- Volume 28, CCIS 373, HCI International 2013 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 29, CCIS 374, HCI International 2013 Posters Proceedings (Part II), edited by Constantine Stephanidis

I would like to thank the Program Chairs and the members of the Program Boards of all affiliated conferences and thematic areas, listed below, for their contribution to the highest scientific quality and the overall success of the HCI International 2013 conference.

This conference could not have been possible without the continuous support and advice of the Founding Chair and Conference Scientific Advisor, Prof. Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution towards the smooth organization of the HCI International 2013 Conference the members of the Human-Computer Interaction Laboratory of ICS-FORTH, and in particular George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, Maria Bouhli and George Kapnas.

May 2013

Constantine Stephanidis  
General Chair, HCI International 2013

# Organization

## Human–Computer Interaction

### Program Chair: Masaaki Kurosu, Japan

Jose Abdelnour-Nocera, UK	Kyungdoh Kim, South Korea
Sebastiano Bagnara, Italy	Heidi Krömker, Germany
Simone Barbosa, Brazil	Chen Ling, USA
Tomas Berns, Sweden	Yan Liu, USA
Nigel Bevan, UK	Zhengjie Liu, P.R. China
Simone Borsci, UK	Loïc Martínez Normand, Spain
Apala Lahiri Chavan, India	Chang S. Nam, USA
Sherry Chen, Taiwan	Naoko Okuizumi, Japan
Kevin Clark, USA	Noriko Osaka, Japan
Torkil Clemmensen, Denmark	Philippe Palanque, France
Xiaowen Fang, USA	Hans Persson, Sweden
Shin'ichi Fukuzumi, Japan	Ling Rothrock, USA
Vicki Hanson, UK	Naoki Sakakibara, Japan
Ayako Hashizume, Japan	Dominique Scapin, France
Anzai Hiroyuki, Italy	Guangfeng Song, USA
Sheue-Ling Hwang, Taiwan	Sanjay Tripathi, India
Wonil Hwang, South Korea	Chui Yin Wong, Malaysia
Minna Isomursu, Finland	Toshiki Yamaoka, Japan
Yong Gu Ji, South Korea	Kazuhiko Yamazaki, Japan
Esther Jun, USA	Ryoji Yoshitake, Japan
Mitsuhiko Karashima, Japan	Silvia Zimmermann, Switzerland

## Human Interface and the Management of Information

### Program Chair: Sakae Yamamoto, Japan

Hans-Jorg Bullinger, Germany	Mark Lehto, USA
Alan Chan, Hong Kong	Hiroyuki Miki, Japan
Gilsoo Cho, South Korea	Hirohiko Mori, Japan
Jon R. Gunderson, USA	Fiona Fui-Hoon Nah, USA
Shin'ichi Fukuzumi, Japan	Shogo Nishida, Japan
Michitaka Hirose, Japan	Robert Proctor, USA
Jhilmil Jain, USA	Youngho Rhee, South Korea
Yasufumi Kume, Japan	Katsunori Shimohara, Japan

Michale Smith, USA  
 Tsutomu Tabe, Japan  
 Hiroshi Tsuji, Japan

Kim-Phuong Vu, USA  
 Tomio Watanabe, Japan  
 Hidekazu Yoshikawa, Japan

## Engineering Psychology and Cognitive Ergonomics

### Program Chair: Don Harris, UK

Guy Andre Boy, USA  
 Joakim Dahlman, Sweden  
 Trevor Dobbins, UK  
 Mike Feary, USA  
 Shan Fu, P.R. China  
 Michaela Heese, Austria  
 Hung-Sying Jing, Taiwan  
 Wen-Chin Li, Taiwan  
 Mark A. Neerinx, The Netherlands  
 Jan M. Noyes, UK  
 Taezoon Park, Singapore

Paul Salmon, Australia  
 Axel Schulte, Germany  
 Siraj Shaikh, UK  
 Sarah C. Sharples, UK  
 Anthony Smoker, UK  
 Neville A. Stanton, UK  
 Alex Stedmon, UK  
 Xianghong Sun, P.R. China  
 Andrew Thatcher, South Africa  
 Matthew J.W. Thomas, Australia  
 Rolf Zon, The Netherlands

## Universal Access in Human–Computer Interaction

### Program Chairs: Constantine Stephanidis, Greece, and Margherita Antona, Greece

Julio Abascal, Spain  
 Ray Adams, UK  
 Gisela Susanne Bahr, USA  
 Margit Betke, USA  
 Christian Bühler, Germany  
 Stefan Carmien, Spain  
 Jerzy Charytonowicz, Poland  
 Carlos Duarte, Portugal  
 Pier Luigi Emiliani, Italy  
 Qin Gao, P.R. China  
 Andrina Granić, Croatia  
 Andreas Holzinger, Austria  
 Josette Jones, USA  
 Simeon Keates, UK

Georgios Kouroupetroglou, Greece  
 Patrick Langdon, UK  
 Seongil Lee, Korea  
 Ana Isabel B.B. Paraguay, Brazil  
 Helen Petrie, UK  
 Michael Pieper, Germany  
 Enrico Pontelli, USA  
 Jaime Sanchez, Chile  
 Anthony Savidis, Greece  
 Christian Stary, Austria  
 Hirotada Ueda, Japan  
 Gerhard Weber, Germany  
 Harald Weber, Germany

## Virtual, Augmented and Mixed Reality

### Program Chair: Randall Shumaker, USA

Waymon Armstrong, USA  
 Juan Cendan, USA  
 Rudy Darken, USA  
 Cali M. Fidopiastis, USA  
 Charles Hughes, USA  
 David Kaber, USA  
 Hirokazu Kato, Japan  
 Denis Laurendeau, Canada  
 Fotis Liarokapis, UK

Mark Livingston, USA  
 Michael Macedonia, USA  
 Gordon Mair, UK  
 Jose San Martin, Spain  
 Jacquelyn Morie, USA  
 Albert “Skip” Rizzo, USA  
 Kay Stanney, USA  
 Christopher Stapleton, USA  
 Gregory Welch, USA

## Cross-Cultural Design

### Program Chair: P.L. Patrick Rau, P.R. China

Pilsung Choe, P.R. China  
 Henry Been-Lirn Duh, Singapore  
 Vanessa Evers, The Netherlands  
 Paul Fu, USA  
 Zhiyong Fu, P.R. China  
 Fu Guo, P.R. China  
 Sung H. Han, Korea  
 Toshikazu Kato, Japan  
 Dyi-Yih Michael Lin, Taiwan  
 Rungtai Lin, Taiwan

Sheau-Farn Max Liang, Taiwan  
 Liang Ma, P.R. China  
 Alexander Mädche, Germany  
 Katsuhiko Ogawa, Japan  
 Tom Plocher, USA  
 Kerstin Röse, Germany  
 Supriya Singh, Australia  
 Hsiu-Ping Yueh, Taiwan  
 Liang (Leon) Zeng, USA  
 Chen Zhao, USA

## Online Communities and Social Computing

### Program Chairs: A. Ant Ozok, USA, and Panayiotis Zaphiris, Cyprus

Areej Al-Wabil, Saudi Arabia  
 Leonelo Almeida, Brazil  
 Bjørn Andersen, Norway  
 Chee Siang Ang, UK  
 Aneesha Bakharia, Australia  
 Ania Bobrowicz, UK  
 Paul Cairns, UK  
 Farzin Deravi, UK  
 Andri Ioannou, Cyprus  
 Slava Kisilevich, Germany

Niki Lambropoulos, Greece  
 Effie Law, Switzerland  
 Soo Ling Lim, UK  
 Fernando Loizides, Cyprus  
 Gabriele Meiselwitz, USA  
 Anthony Norcio, USA  
 Elaine Raybourn, USA  
 Panote Siriaraya, UK  
 David Stuart, UK  
 June Wei, USA

## **Augmented Cognition**

### **Program Chairs: Dylan D. Schmorrow, USA, and Cali M. Fidopiastis, USA**

Robert Arrabito, Canada

Richard Backs, USA

Chris Berka, USA

Joseph Cohn, USA

Martha E. Crosby, USA

Julie Drexler, USA

Ivy Estabrooke, USA

Chris Forsythe, USA

Wai Tat Fu, USA

Rodolphe Gentili, USA

Marc Grootjen, The Netherlands

Jefferson Grubb, USA

Ming Hou, Canada

Santosh Mathan, USA

Rob Matthews, Australia

Dennis McBride, USA

Jeff Morrison, USA

Mark A. Neerincx, The Netherlands

Denise Nicholson, USA

Banu Onaral, USA

Lee Sciarini, USA

Kay Stanney, USA

Roy Stripling, USA

Rob Taylor, UK

Karl van Orden, USA

## **Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management**

### **Program Chair: Vincent G. Duffy, USA and Russia**

Karim Abdel-Malek, USA

Giuseppe Andreoni, Italy

Daniel Carruth, USA

Eliza Yingzi Du, USA

Enda Fallon, Ireland

Afzal Godil, USA

Ravindra Goonetilleke, Hong Kong

Bo Hoege, Germany

Waldemar Karwowski, USA

Zhizhong Li, P.R. China

Kang Li, USA

Tim Marler, USA

Michelle Robertson, USA

Matthias Rötting, Germany

Peter Vink, The Netherlands

Mao-Jiun Wang, Taiwan

Xuguang Wang, France

Jingzhou (James) Yang, USA

Xiugan Yuan, P.R. China

Gülcin Yücel Hoge, Germany

## **Design, User Experience, and Usability**

### **Program Chair: Aaron Marcus, USA**

Sisira Adikari, Australia

Ronald Baecker, Canada

Arne Berger, Germany

Jamie Blustein, Canada

Ana Boa-Ventura, USA

Jan Brejcha, Czech Republic

Lorenzo Cantoni, Switzerland

Maximilian Eibl, Germany

Anthony Faiola, USA  
 Emilie Gould, USA  
 Zelda Harrison, USA  
 Rüdiger Heimgärtner, Germany  
 Brigitte Herrmann, Germany  
 Steffen Hess, Germany  
 Kaleem Khan, Canada

Jennifer McGinn, USA  
 Francisco Rebelo, Portugal  
 Michael Renner, Switzerland  
 Kerem Rızvanoğlu, Turkey  
 Marcelo Soares, Brazil  
 Christian Sturm, Germany  
 Michele Visciola, Italy

## **Distributed, Ambient and Pervasive Interactions**

### **Program Chairs: Norbert Streitz, Germany, and Constantine Stephanidis, Greece**

Emile Aarts, The Netherlands  
 Adnan Abu-Dayya, Qatar  
 Juan Carlos Augusto, UK  
 Boris de Ruyter, The Netherlands  
 Anind Dey, USA  
 Dimitris Grammenos, Greece  
 Nuno M. Guimaraes, Portugal  
 Shin'ichi Konomi, Japan  
 Carsten Magerkurth, Switzerland

Christian Müller-Tomfelde, Australia  
 Fabio Paternó, Italy  
 Gilles Privat, France  
 Harald Reiterer, Germany  
 Carsten Röcker, Germany  
 Reiner Wichert, Germany  
 Woontack Woo, South Korea  
 Xenophon Zabulis, Greece

## **Human Aspects of Information Security, Privacy and Trust**

### **Program Chairs: Louis Marinou, ENISA EU, and Ioannis Askoxylakis, Greece**

Claudio Agostino Ardagna, Italy  
 Zinaida Benenson, Germany  
 Daniele Catteddu, Italy  
 Raoul Chiesa, Italy  
 Bryan Cline, USA  
 Sadie Creese, UK  
 Jorge Cuellar, Germany  
 Marc Dacier, USA  
 Dieter Gollmann, Germany  
 Kirstie Hawkey, Canada  
 Jaap-Henk Hoepman, The Netherlands  
 Cagatay Karabat, Turkey  
 Angelos Keromytis, USA  
 Ayako Komatsu, Japan

Ronald Leenes, The Netherlands  
 Javier Lopez, Spain  
 Steve Marsh, Canada  
 Gregorio Martinez, Spain  
 Emilio Mordini, Italy  
 Yuko Murayama, Japan  
 Masakatsu Nishigaki, Japan  
 Aljosa Pasic, Spain  
 Milan Petković, The Netherlands  
 Joachim Posegga, Germany  
 Jean-Jacques Quisquater, Belgium  
 Damien Sauveron, France  
 George Spanoudakis, UK  
 Kerry-Lynn Thomson, South Africa



Julien Touzeau, France  
Theo Tryfonas, UK  
João Vilela, Portugal

Claire Vishik, UK  
Melanie Volkamer, Germany

## External Reviewers

Maysoon Abulhair, Saudi Arabia  
Ilia Adami, Greece  
Vishal Barot, UK  
Stephan Böhm, Germany  
Vassilis Charissis, UK  
Francisco Cipolla-Ficarra, Spain  
Maria De Marsico, Italy  
Marc Fabri, UK  
David Fonseca, Spain  
Linda Harley, USA  
Yasushi Ikei, Japan  
Wei Ji, USA  
Nouf Khashman, Canada  
John Killilea, USA  
Iosif Klironomos, Greece  
Ute Klotz, Switzerland  
Maria Korozi, Greece  
Kentaro Kotani, Japan

Vassilis Kouroumalis, Greece  
Stephanie Lackey, USA  
Janelle LaMarche, USA  
Asterios Leonidis, Greece  
Nickolas Macchiarella, USA  
George Margetis, Greece  
Matthew Marraffino, USA  
Joseph Mercado, USA  
Claudia Mont'Alvão, Brazil  
Yoichi Motomura, Japan  
Karsten Nebe, Germany  
Stavroula Ntoa, Greece  
Martin Osen, Austria  
Stephen Prior, UK  
Farid Shirazi, Canada  
Jan Stelovsky, USA  
Sarah Swierenga, USA

# HCI International 2014

The 16th International Conference on Human–Computer Interaction, HCI International 2014, will be held jointly with the affiliated conferences in the summer of 2014. It will cover a broad spectrum of themes related to Human–Computer Interaction, including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: <http://www.hci-international.org/>

General Chair

Professor Constantine Stephanidis  
University of Crete and ICS-FORTH  
Heraklion, Crete, Greece  
Email: [cs@ics.forth.gr](mailto:cs@ics.forth.gr)

# Table of Contents – Part I

## Developing Augmented and Virtual Environments

Passive Viewpoints in a Collaborative Immersive Environment . . . . .	3
<i>Sarah Coburn, Lisa Rebenitsch, and Charles Owen</i>	
Virtual Reality Based Interactive Conceptual Simulations: Combining Post-processing and Linear Static Simulations . . . . .	13
<i>Holger Graf and André Stork</i>	
Enhancing Metric Perception with RGB-D Camera . . . . .	23
<i>Daiki Handa, Hirotake Ishii, and Hiroshi Shimoda</i>	
Painting Alive: Handheld Augmented Reality System for Large Targets . . . . .	32
<i>Jae-In Hwang, Min-Hyuk Sung, Ig-Jae Kim, Sang Chul Ahn, Hyoung-Gon Kim, and Heedong Ko</i>	
VWSocialLab: Prototype Virtual World (VW) Toolkit for Social and Behavioral Science Experimental Set-Up and Control . . . . .	39
<i>Lana Jaff, Austen Hayes, and Amy Ulinski Banic</i>	
Controlling and Filtering Information Density with Spatial Interaction Techniques via Handheld Augmented Reality . . . . .	49
<i>Jens Keil, Michael Zoellner, Timo Engelke, Folker Wientapper, and Michael Schmitt</i>	
Development of Multiview Image Generation Simulator for Depth Map Quantization . . . . .	58
<i>Minyoung Kim, Ki-Young Seo, Seokhwan Kim, Kyoung Shin Park, and Yongjoo Cho</i>	
Authoring System Using Panoramas of Real World . . . . .	65
<i>Hee Jae Kim and Jong Weon Lee</i>	
Integrated Platform for an Augmented Environment with Heterogeneous Multimodal Displays . . . . .	73
<i>Jaedong Lee, Sangyong Lee, and Gerard Jounghyun Kim</i>	
Optimal Design of a Haptic Device for a Particular Task in a Virtual Environment . . . . .	79
<i>Jose San Martin, Loic Corenthy, Luis Pastor, and Marcos Garcia</i>	
Real-Time Dynamic Lighting Control of an AR Model Based on a Data-Glove with Accelerometers and NI-DAQ . . . . .	86
<i>Alex Rodiera Clarens and Isidro Navarro</i>	

Ultra Low Cost Eye Gaze Tracking for Virtual Environments . . . . .	94
<i>Matthew Swarts and Jin Noh</i>	
Real-Time Stereo Rendering Technique for Virtual Reality System Based on the Interactions with Human View and Hand Gestures . . . . .	103
<i>Viet Tran Hoang, Anh Nguyen Hoang, and Dongho Kim</i>	
Information Management for Multiple Entities in a Remote Sensor Environment . . . . .	111
<i>Peter Venero, Allen Rowe, Thomas Carretta, and James Boyer</i>	

**Interaction in Augmented and Virtual Environments**

Tactile Apparent Motion Presented from Seat Pan Facilitates Racing Experience . . . . .	121
<i>Tomohiro Amemiya, Koichi Hirota, and Yasushi Ikei</i>	
Predicting Navigation Performance with Psychophysiological Responses to Threat in a Virtual Environment . . . . .	129
<i>Christopher G. Courtney, Michael E. Dawson, Albert A. Rizzo, Brian J. Arizmendi, and Thomas D. Parsons</i>	
A Study of Navigation and Selection Techniques in Virtual Environments Using Microsoft Kinect® . . . . .	139
<i>Peter Dam, Priscilla Braz, and Alberto Raposo</i>	
Legibility of Letters in Reality, 2D and 3D Projection . . . . .	149
<i>Elisabeth Dittrich, Stefan Brandenburg, and Boris Beckmann-Dobrev</i>	
The Visual, the Auditory and the Haptic – A User Study on Combining Modalities in Virtual Worlds . . . . .	159
<i>Julia Fröhlich and Ipke Wachsmuth</i>	
Spatial Augmented Reality on Person: Exploring the Most Personal Medium . . . . .	169
<i>Adrian S. Johnson and Yu Sun</i>	
Parameter Comparison of Assessing Visual Fatigue Induced by Stereoscopic Video Services . . . . .	175
<i>Kimiko Kawashima, Jun Okamoto, Kazuo Ishikawa, and Kazuno Negishi</i>	
Human Adaptation, Plasticity and Learning for a New Sensory-Motor World in Virtual Reality . . . . .	184
<i>Michiteru Kitazaki</i>	
An Asymmetric Bimanual Gestural Interface for Immersive Virtual Environments . . . . .	192
<i>Julien-Charles Lévesque, Denis Laurendeau, and Marielle Mokhtari</i>	

A New Approach for Indoor Navigation Using Semantic Webtechnologies and Augmented Reality . . . . .	202
<i>Tamás Matuszka, Gergő Gombos, and Attila Kiss</i>	
Assessing Engagement in Simulation-Based Training Systems for Virtual Kinesic Cue Detection Training . . . . .	211
<i>Eric Ortiz, Crystal Maraj, Julie Salcedo, Stephanie Lackey, and Irwin Hudson</i>	
Development of Knife-Shaped Interaction Device Providing Virtual Tactile Sensation . . . . .	221
<i>Azusa Toda, Kazuki Tanaka, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura</i>	
GUI Design Solution for a Monocular, See-through Head-Mounted Display Based on Users' Eye Movement Characteristics . . . . .	231
<i>Takahiro Uchiyama, Kazuhiro Tanuma, Yusuke Fukuda, and Miwa Nakanishi</i>	
Visual, Vibrotactile, and Force Feedback of Collisions in Virtual Environments: Effects on Performance, Mental Workload and Spatial Orientation . . . . .	241
<i>Bernhard Weber, Mikel Sagardia, Thomas Hulin, and Carsten Preusche</i>	

## **Human-Robot Interaction in Virtual Environments**

What Will You Do Next? A Cognitive Model for Understanding Others' Intentions Based on Shared Representations . . . . .	253
<i>Haris Dindo and Antonio Chella</i>	
Toward Task-Based Mental Models of Human-Robot Teaming: A Bayesian Approach . . . . .	267
<i>Michael A. Goodrich and Daqing Yi</i>	
Assessing Interfaces Supporting Situational Awareness in Multi-agent Command and Control Tasks . . . . .	277
<i>Donald Kalar and Collin Green</i>	
Cognitive Models of Decision Making Processes for Human-Robot Interaction . . . . .	285
<i>Christian Lebiere, Florian Jentsch, and Scott Ososky</i>	
Human Considerations in the Application of Cognitive Decision Models for HRI . . . . .	295
<i>Scott Ososky, Florian Jentsch, and Elizabeth Phillips</i>	

Computational Mechanisms for Mental Models in Human-Robot Interaction . . . . .	304
<i>Matthias Scheutz</i>	
Increasing Robot Autonomy Effectively Using the Science of Teams . . . .	313
<i>David Schuster and Florian Jentsch</i>	
Cybernetic Teams: Towards the Implementation of Team Heuristics in HRI . . . . .	321
<i>Travis J. Wiltshire, Dustin C. Smith, and Joseph R. Keebler</i>	
<b>Presence and Tele-presence</b>	
Embodiment and Embodied Cognition . . . . .	333
<i>Mark R. Costa, Sung Yeun Kim, and Frank Biocca</i>	
DigiLog Space Generator for Tele-Collaboration in an Augmented Reality Environment . . . . .	343
<i>Kyungwon Gil, Taejin Ha, and Woontack Woo</i>	
Onomatopoeia Expressions for Intuitive Understanding of Remote Office Situation . . . . .	351
<i>Kyota Higa, Masumi Ishikawa, and Toshiyuki Nomura</i>	
Enhancing Social Presence in Augmented Reality-Based Telecommunication System . . . . .	359
<i>Jea In Kim, Taejin Ha, Woontack Woo, and Chung-Kon Shi</i>	
How Fiction Informed the Development of Telepresence and Teleoperation . . . . .	368
<i>Gordon M. Mair</i>	
High Presence Communication between the Earth and International Space Station . . . . .	378
<i>Tetsuro Ogi, Yoshisuke Tateyama, and Yosuke Kubota</i>	
Effects of Visual Fidelity on Biometric Cue Detection in Virtual Combat Profiling Training . . . . .	388
<i>Julie Salcedo, Crystal Maraj, Stephanie Lackey, Eric Ortiz, Irwin Hudson, and Joy Martinez</i>	
<b>Author Index</b> . . . . .	397

## Table of Contents – Part II

### Healthcare and Medical Applications

Gait Analysis Management and Diagnosis in a Prototype Virtual Reality Environment . . . . .	3
<i>Salsabeel F.M. Alfalah, David K. Harrison, and Vassilis Charissis</i>	
Theory-Guided Virtual Reality Psychotherapies: Going beyond CBT-Based Approaches . . . . .	12
<i>Sheryl Brahmam</i>	
Development of the Home Arm Movement Stroke Training Environment for Rehabilitation (HAMSTER) and Evaluation by Clinicians . . . . .	22
<i>Elizabeth B. Brokaw and Bambi R. Brewer</i>	
A Low Cost Virtual Reality System for Rehabilitation of Upper Limb . . . . .	32
<i>Pawel Budziszewski</i>	
Super Pop VR™: An Adaptable Virtual Reality Game for Upper-Body Rehabilitation . . . . .	40
<i>Sergio García-Vergara, Yu-Ping Chen, and Ayanna M. Howard</i>	
Asynchronous Telemedicine Diagnosis of Musculoskeletal Injuries through a Prototype Interface in Virtual Reality Environment . . . . .	50
<i>Soheeb Khan, Vassilis Charissis, David K. Harrison, Sophia Sakellariou, and Warren Chan</i>	
Developing a Theory-Informed Interactive Animation to Increase Physical Activity among Young People with Asthma . . . . .	60
<i>Jennifer Murray, Brian Williams, Gaylor Hoskins, John McGhee, Dylan Gauld, and Gordon Brown</i>	
The Design Considerations of a Virtual Reality Application for Heart Anatomy and Pathology Education . . . . .	66
<i>Victor Nyamse, Vassilis Charissis, J. David Moore, Caroline Parker, Soheeb Khan, and Warren Chan</i>	
Human-Computer Confluence for Rehabilitation Purposes after Stroke . . . . .	74
<i>Rupert Ortner, David Ram, Alexander Kollreider, Harald Pitsch, Joanna Wojtowicz, and Günter Edlinger</i>	
Projected AR-Based Interactive CPR Simulator . . . . .	83
<i>Nohyoung Park, Yeram Kwon, Sungwon Lee, Woontack Woo, and Jihoon Jeong</i>	

Affecting Our Perception of Satiety by Changing the Size of Virtual Dishes Displayed with a Tabletop Display . . . . .	90
<i>Sho Sakurai, Takuji Narumi, Yuki Ban, Tomohiro Tanikawa, and Michitaka Hirose</i>	

## Virtual and Augmented Environments for Learning and Education

An Experience on Natural Sciences Augmented Reality Contents for Preschoolers . . . . .	103
<i>Antonia Cascales, Isabel Laguna, David Pérez-López, Pascual Perona, and Manuel Contero</i>	
Teaching 3D Arts Using Game Engines for Engineering and Architecture . . . . .	113
<i>Jaume Duran and Sergi Villagrasa</i>	
The Characterisation of a Virtual Reality System to Improve the Quality and to Reduce the Gap between Information Technology and Medical Education . . . . .	122
<i>Jannat Falah, David K. Harrison, Vassilis Charissis, and Bruce M. Wood</i>	
A Mobile Personal Learning Environment Approach . . . . .	132
<i>Francisco José García-Peñalvo, Miguel Ángel Conde, and Alberto Del Pozo</i>	
Perceived Presence's Role on Learning Outcomes in a Mixed Reality Classroom of Simulated Students . . . . .	142
<i>Aleshia T. Hayes, Stacey E. Hardin, and Charles E. Hughes</i>	
The Building as the Interface: Architectural Design for Education in Virtual Worlds . . . . .	152
<i>Luis Antonio Hernández Ibáñez and Viviana Barneche Naya</i>	
Mixed Reality Space Travel for Physics Learning . . . . .	162
<i>Darin E. Hughes, Shabnam Sabbagh, Robb Lindgren, J. Michael Moshell, and Charles E. Hughes</i>	
Picking Up STEAM: Educational Implications for Teaching with an Augmented Reality Guitar Learning System . . . . .	170
<i>Joseph R. Keebler, Travis J. Wiltshire, Dustin C. Smith, and Stephen M. Fiore</i>	
Virtual Reality Data Visualization for Team-Based STEAM Education: Tools, Methods, and Lessons Learned . . . . .	179
<i>Daniel F. Keefe and David H. Laidlaw</i>	



Architectural Geo-E-Learning: Geolocated Teaching in Urban Environments with Mobile Devices. A Case Study and Work In Progress . . . . .	188
<i>Ernest Redondo, Albert Sánchez Riera, David Fonseca, and Alberto Peredo</i>	

## **Business, Industrial and Military Applications**

Mixed Reality Environment for Mission Critical Systems Servicing and Repair . . . . .	201
<i>Andrea F. Abate, Fabio Narducci, and Stefano Ricciardi</i>	
Establishing Workload Manipulations Utilizing a Simulated Environment . . . . .	211
<i>Julian Abich IV, Lauren Reinerman-Jones, and Grant Taylor</i>	
Interactive Virtual Reality Shopping and the Impact in Luxury Brands . . . . .	221
<i>Samar Altarteer, Vassilis Charissis, David Harrison, and Warren Chan</i>	
Multiple Remotely Piloted Aircraft Control: Visualization and Control of Future Path . . . . .	231
<i>Gloria Calhoun, Heath Ruff, Chad Breeden, Joshua Hamell, Mark Draper, and Christopher Miller</i>	
The Virtual Dressing Room: A Perspective on Recent Developments . . . .	241
<i>Michael B. Holte</i>	
Making Sense of Large Datasets in the Context of Complex Situation Understanding . . . . .	251
<i>Marielle Mokhtari, Eric Boivin, and Denis Laurendeau</i>	
Evaluating Distraction and Disengagement of Attention from the Road . . . . .	261
<i>Valentine Nwakacha, Andy Crabtree, and Gary Burnett</i>	
DCS 3D Operators in Industrial Environments: New HCI Paradigm for the Industry . . . . .	271
<i>Manuel Pérez-Cota and Miguel Ramón González Castro</i>	
Natural Feature Tracking Augmented Reality for On-Site Assembly Assistance Systems . . . . .	281
<i>Rafael Radkowski and James Oliver</i>	
Augmented Reality Interactive System to Support Space Planning Activities . . . . .	291
<i>Guido Maria Re, Giandomenico Caruso, and Monica Bordegoni</i>	

Empirical Investigation of Transferring Cockpit Interactions from Virtual to Real-Life Environments . . . . .	301
<i>Diana Reich and Elisabeth Dittrich</i>	
Mixed and Augmented Reality for Marine Corps Training . . . . .	310
<i>Richard Schaffer, Sean Cullen, Phe Meas, and Kevin Dill</i>	
Proactive Supervisory Decision Support from Trend-Based Monitoring of Autonomous and Automated Systems: A Tale of Two Domains . . . . .	320
<i>Harvey S. Smallman and Maia B. Cook</i>	
The ART of CSI: An Augmented Reality Tool (ART) to Annotate Crime Scenes in Forensic Investigation . . . . .	330
<i>Jan Willem Streefkerk, Mark Houben, Pjotr van Amerongen, Frank ter Haar, and Judith Dijk</i>	
The Virtual Reality Applied in Construction Machinery Industry . . . . .	340
<i>Yun-feng Wu, Ying Zhang, Jun-wu Shen, and Tao Peng</i>	

**Culture and Entertainment Applications**

On the Use of Augmented Reality Technology for Creating Interactive Computer Games . . . . .	353
<i>Chin-Shyurng Fahn, Meng-Luen Wu, and Wei-Tyng Liu</i>	
A 3-D Serious Game to Simulate a Living of a Beehive . . . . .	363
<i>José Eduardo M. de Figueiredo, Vera Maria B. Werneck, and Rosa Maria E. Moreira da Costa</i>	
Presentation of Odor in Multi-Sensory Theater . . . . .	372
<i>Koichi Hirota, Yoko Ito, Tomohiro Amemiya, and Yasushi Ikei</i>	
Using Motion Sensing for Learning: A Serious Nutrition Game . . . . .	380
<i>Mina C. Johnson-Glenberg</i>	
AR’istophanes: Mixed Reality Live Stage Entertainment with Spectator Interaction . . . . .	390
<i>Thiemo Kastel, Marion Kesmaecker, Krzysztof Mikolajczyk, and Bruno Filipe Duarte-Gonçalves</i>	
System Development of Immersive Technology Theatre in Museum . . . . .	400
<i>Yi-Chia Nina Lee, Li-Ting Shan, and Chien-Hsu Chen</i>	
An Immersive Environment for a Virtual Cultural Festival . . . . .	409
<i>Liang Li, Woong Choi, and Kozaburo Hachimura</i>	
Mission: LEAP Teaching Innovation Competencies by Mixing Realities . . . . .	416
<i>Christopher Stapleton, Atsusi “2C” Hirumi, and Dana S. Mott</i>	

ChronoLeap: The Great World’s Fair Adventure ..... 426  
*Lori C. Walters, Darin E. Hughes, Manuel Gétrudix Barrio, and  
Charles E. Hughes*

The Electric Bow Interface ..... 436  
*Masasuke Yasumoto and Takashi Ohta*

**Author Index** ..... 443

## **Part I**

# **Developing Augmented and Virtual Environments**

# Passive Viewpoints in a Collaborative Immersive Environment

Sarah Coburn, Lisa Rebenitsch, and Charles Owen

Computer Science & Engineering, Michigan State University, East Lansing, MI, USA  
{coburnsa, rebenits, cbowen}@cse.msu.edu

**Abstract.** Widespread acceptance of virtual reality has been partially handicapped by the inability of current systems to accommodate multiple viewpoints, thereby limiting their appeal for collaborative applications. We are exploring the ability to utilize passive, untracked participants in a powerwall environment. These participants see the same image as the active, immersive participant. This does present the passive user with a varying viewpoint that does not correspond to their current position. We demonstrate the impact this will have on the perceived image and show that human psychology is actually well adapted to compensating for what, on the surface, would seem to be a very drastic distortion. We present some initial guidelines for system design that minimize the negative impact of passive participation, allowing two or more collaborative participants. We then outline future experimentation to measure user compensation for these distorted viewpoints.

**Keywords:** virtual reality, passive participation, immersion, perception.

## 1 Introduction

Many immersive environments are unavailable to the general public because they are limited to large rooms and they require a tremendous budget (sometimes even exceeding \$1 million [1]). One solution is the one-walled VR environment (the “power-wall”), which can be implemented with a single screen and projector, and can use a number of different widely available tracking systems.

Many existing virtual reality (VR) systems only allow one or two users. When VR demos or experiments are performed, extra shutter or polarized glasses are often used for additional viewers to sit in and watch the world in the role of passive participants in the virtual experience. With this in mind, can we create collaborative VR systems that can actually involve these passive participants, allowing them to experience and interact with the virtual environment, alongside an active, tracked participant? A common obstacle to the adoption of VR is that so many systems are fundamentally solitary experiences. Some hardware approaches allow for even up to six active users, but these approaches are expensive, difficult to set up and calibrate, and not as available as basic single-user systems. In order to display individual perspectives, these systems often add multiple projectors and polarized images in order to project images to the left and right eyes of each participant [2][3][4]. These systems show promise

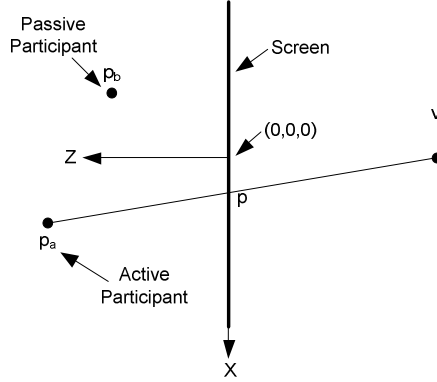
for the next generation of VR collaboration, but the resulting cost and computation power makes it less accessible to the general public, as well as difficult to extend to even more participants. As a result, we are examining the design of systems that will allow multiple passive users to participate with an active user in a VR environment, possibility with the ability to pass the active experience among users as necessary, so as to support collaborative applications. This paper addresses some of the physical and psychological issues associated with the participation of passive users and explores some initial design guidelines to tailor applications for passive participation.

One of the greatest challenges in designing a collaborative virtual environment is making sure that all participants can accurately perceive the virtual world. In simpler virtual environments, if multiple participants view the world, virtual objects are typically presented based on a single center of projection from an oblique angle. To design a better multi-user, one-walled VR system, the design must have knowledge of where the passive participants are, as well as what elements in the environment can be manipulated to counteract the perception challenges. Knowing the location of the user allows manipulation of the virtual world to actively adapt to passive participants, even when there is still only a solitary center of projection. In addition, the VR system can take advantage of the psychological ability to compensate for distortions.

In this article, we explore the appearance and perception of the virtual experience as seen by a passive participant. Section 2 presents a mathematical model for the image as seen by passive participants, regardless of their relative position to the screen or the active participant. Section 3 explores the psychological and physiological components of a passive participant's perception of the virtual world. Finally, Section 4 establishes some leading guidelines for the design of virtual environments that take advantage of these elements.

## 2 Mathematical Basis

One of the most basic VR systems is the powerwall, which uses a rear projection stereo display or a large 3D monitor to provide a virtual experience for a single user with 3D glasses and head tracking. Fig.1 illustrates the basic projection model for a powerwall. It is assumed that the world coordinate system origin is centered on the screen with the Y axis up and the X axis to the right and we assume a right-hand coordinate system. In the figure, the point  $p_a (x_a, y_a, z_a)$  represents the active participant. This participant is tracked and the basis for the projection. The point  $p_b (x_b, y_b, z_b)$  represents a passive participant. This participant is untracked and sees the same image as the active participant. The point  $v (x, y, z)$  represents a point in the virtual space subject to projection. The point  $p (x_p, y_p, z_p)$  represents the projected point on the screen. Projection is accomplished by creating a *projector*, a ray starting at the center of projection and pointing at the virtual point, and computing the intersection of that ray with the virtual projection surface.



**Fig. 1.** Model for screen-based VR projection

Given this orientation, projection is implemented in real systems by changing the origin to the active participant viewpoint and multiplying by the ratio of the distance of the projection screen (focal length) and the z-axis distance to the point. (Real-world systems also include a mapping from the virtual projection screen to pixels on the real projection screen, but that linear mapping does not affect the results in this paper and is, therefore, ignored). The basic equations for projection are:

$$x_p = (x - x_a) \left( \frac{-z_a}{z - z_a} \right) \quad (1)$$

$$y_p = (y - y_a) \left( \frac{-z_a}{z - z_a} \right) \quad (2)$$

Since projection for x and y is symmetrical, only the x component will be indicated in future equations. The y component can be determined through simple substitutions. All of the equations in this section apply for virtual points on either side of the screen.

## 2.1 Depth

The passive participant would ideally view the world projected properly for them. However, they instead see the world as projected for that active participant. The first question is: what is the equivalent mapping between the two? The multiplicative factor from Equation 1:

$$\left( \frac{-z_a}{z - z_a} \right) \quad (3)$$

determines the foreshortening due to depth. Foreshortening is the tendency of objects under perspective projection to appear smaller or larger due to their distance from the viewer. The foreshortening seen by the passive participant will be different if the participant is not at the same distance from the screen as the active participant. Intuitively, if the passive participant is half the distance to the screen, the amount of foreshortening will be doubled. This is the same as if the depth of the virtual space was scaled by a factor of 0.5, making objects appear to be half as far away as they are.

We can illustrate this by scaling the  $z$  values for the passive participant's foreshortening term by  $z_p/z_a$ :

$$\frac{-z_p}{\frac{z_p}{z_a} - z_p} = \frac{-z_p z_a}{z z_p - z_p z_a} = \frac{-z_a}{z - z_a} \tag{4}$$

What this means is that a passive participant moving closer to the screen will see a world that appears scaled in the  $z$  dimension.

### 2.2 Position

Given the solution for the change of depth for perceived point, we can now examine the  $x, y$  components of the perceived point for the passive participant. Figure 2 illustrates the geometry of the problem. The point  $v$  is projected to point  $p$  on the screen for the active participant. The passive participant sees that point at point  $p$  as some point along the projector from  $p_b$  through point  $p$  to the point  $v'$  as it appears to that participant. We know the depth of  $v'$  is  $z^{z_p/z_a}$  due to Equation 4.

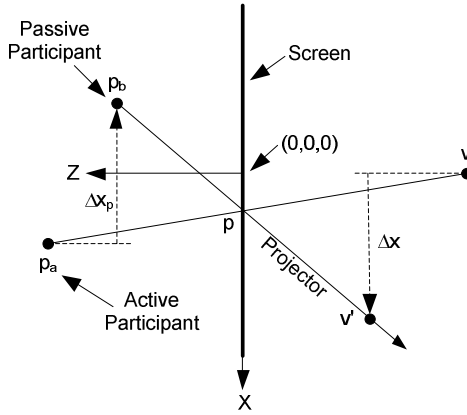


Fig. 2. Geometry of virtual point displacement

The geometry is proportional between the two sides. If the passive participant position is offset from the active participant by  $\Delta x_p = x_p - x_a$ , the perceived position of the point will be offset by the negative of that value scaled by the ratio of the point depth divided by the view distance for the active participant:

$$x' = x + (x_p - x_a) \frac{z}{z_a} \tag{5}$$

Effectively, this imparts a skew on the virtual world. The amount of the skew is determined by the difference between the positions of the active and passive participants and is based on depth. Given these solutions, the passive participant will perceive the point  $(x, y, z)$  at this position:



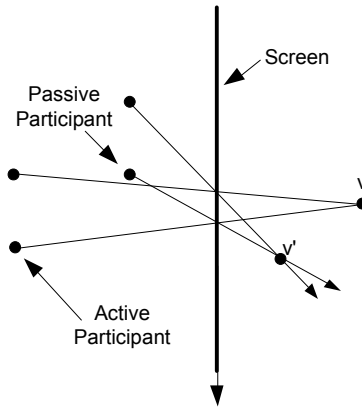
$$x' = x + (x_p - x_a) \frac{z}{z_a} \quad (6)$$

$$y' = y + (y_p - y_a) \frac{z}{z_a} \quad (7)$$

$$z' = z \frac{z_p}{z_a} \quad (8)$$

### 2.3 Stereo

The depth scaling also holds for stereo disparity, assuming the interpupillary distance (IPD) and eye orientations for the active and passive participants are the same. This relationship is illustrated in Figure 3.



**Fig. 3.** Depth scaling due to stereo disparity

As many stereo installations assume only head position tracking and fixed IPD values, the stereo depth scaling will correspond to that due to this varying participant screen distances. If the IPD distance or orientation is allowed to vary, there will be a divergence, which can be confusing.

## 3 Psychological Considerations

As illustrated in Section 2, the distortions resulting from a non-primary viewpoint present the passive observer with an image of object that are skewed and distorted, rather than the correct proportions. However, there is evidence for psychological compensation for this distortion, allowing observers to correctly perceive object proportions, orientations, and relative positions in a variety of visual mediums. From paintings to photographs, people are actually quite accustomed to viewing images from a point that does not correspond to the center of projection.

### 3.1 Potential Psychological Complications

**Judging Object Orientations.** One characteristic of a virtual scene is the orientation of the projected objects. If viewed differently by the individual participants, there is a potential confound of collaboration.

Goldstein [5,6] discussed three main pictorial perception attributes: perceived orientation (in relation to the observer), perceived spatial layout (of the objects in the image), and perceived projection (perception on the observer's retina). He determined that relative position of objects in an image (in the virtual space) was maintained, despite a variety of oblique viewing angles. However, he did find that perceived orientation was not maintained as the difference in angle from the center of projection increased. As a result, Goldstein identified the differential rotation effect (DRE), where objects in a 2D image pointing to the observer (at an angle near  $90^\circ$ ) appear to move faster than objects farther away when the observer moves relative to the image.

One popular virtual collaboration medium is the interactive tabletop. Multiple users interact with a single table, which has virtual objects presented from a common viewpoint. Hancock et al [7] analyzed user perception of orientation of objects projected on a tabletop device, and found that users had a more difficult time judging object orientation when the COP was farther away from their own viewpoint. They determined that a central, neutral COP helped to minimize discrepancies between user perceptions. This study also found that using an orthographic projection (such as found in blueprint drawings to maintain correct object proportions) aided the users in identifying the orientation of objects, though this projection does eliminate foreshortening, an important depth cue. However, though a tabletop design often requires a  $360^\circ$  of potential user viewpoints, CAVE environments (both single and multiple screen displays) naturally limit the range of these viewpoints. This suggests that a neutral viewpoint might be more effective in a CAVE environment than the study found for the tabletop.

**Cybersickness.** A common element in virtual systems is cybersickness, a form of motion sickness associated with virtual environments. While the exact causes of cybersickness are still under investigation, it is known to be common in situations where the camera viewpoint moves independently of movement from passive observers.

Cybersickness symptoms are prevalent in most VR environments and can impact from ~30% [8] to ~80% [9] of participants. Displaying imagery that is not controlled by the observer generally increases symptoms. Swaying imagery is often used in cybersickness research to invoke symptoms. Chen et al. [8] and Dong and Stoffregen [10] demonstrated this effect with experiments where participants were "drivers" or "passengers" in a virtual car. Drivers were given a tracked viewpoint and control of their environment and the resulting video was recorded. This video was then displayed to a passenger. Not surprisingly, passengers had more cybersickness symptoms than drivers. Passengers also had high amounts of body movements, indicating a strong involvement (and conflict) with the movement in the recording.

In both of these situations, participants who suffered cybersickness symptoms were not in control of the camera viewpoint. If a collaborative environment must share a common viewpoint, the design should consider ways to minimize the cybersickness effects that will result. Another major source of cybersickness symptoms is when visual and kinetic information conflict. This makes sharing a common viewpoint difficult: passive viewers are not in control of changes to the perspective.

### 3.2 Compensation for Distortion

In order to determine what characteristics of the virtual scene can be or even need to be modified to help accentuate depth and spatial cues, we must first explore what the brain will use in order to compensate for missing stimuli.

Gestalt psychology here argued that the mind has the ability to organize visual stimuli, and is structured and operates in such a way to make this possible. As opposed to the empirical belief (which says we only learn through experience with the world), Gestalt psychology says that the “brain uses organizational principles to interpret visual appearance” [11].

Because of this ability of the mind to natively organize incoming visual stimuli, the mind can still form a cognitive model of the object in a distorted picture. Since it is not purely empirical, the distortion does not negate the ability to form and access mental models.

**Spatial Cognition.** Spatial cognition is the ability of the brain to acquire, organize, utilize, and revise knowledge about spatial environments [12]. Knowledge of this cognitive ability is especially helpful in creating virtual environments. If certain characteristics of the system hamper the participant’s ability to correctly perceive the virtual world, knowledge of the mind’s organizational power can help compensate for the shortcomings.

The neuroscientist David Marr identified three stages of visual processing, including what he coined the “3D sketch,” when the mind visualizes the world in a mental 3D model [13]. Intelligent perception allows the mind to make cognitive observations about this 3D data. One theory of the brain’s ability to categorize visual information is that rather than purely learning from experience (empirical view) the brain actively organizes incoming visual sensory data into mental models. Gestalt psychology established this theory, and began to identify the specific characteristics that the brain uses to categorize visual data.

People routinely view 2D imagery from a wide range of angles and have little difficulty with correctly interpreting the data in the image. Vishwanath et al. [14] suggested that 2D pictures look correct when viewed from the wrong angle not by geometric data inside the picture itself, but from an estimate of local surface orientation. When binocular vision was available, participants were able to correctly judge object characteristics invariant of their viewing angle. Instead of judging object characteristics by observing the objects within the image, participants instead used cues from the local slant of the image itself to correct perspective distortion from viewing the picture obliquely.

**Spatial and Depth Cues.** Despite the known variations of object characteristics between different viewpoints, there are many other cues that the human brain will use to determine spatial characteristics. As discussed by Christou and Parker [11], there are several pictorial cues that allow 2D images to represent and simulate 3D objects. These cues can be manipulated by a virtual system, and can simulate natural objects by imitating perspective and depth. Most of these cues can be emphasized to compensate for any existing distortions:

*Shading:* Shading and gradients can easily be used to make closer elements brighter and clearer, while more distant objects (or the distant parts of an object) will have darker or dimmer shading, depending on the lighting in the scene.

*Color Fading:* Similar to shading, color fades when farther away, so closer objects will have higher color saturation. Atmospheric effects also add to this: objects farther in the distance are often dimmed due to water, dust, or other particles in the air.

*Interposition:* Objects closer to the observers will occlude objects farther away.

*Shadows:* Shadows from the objects in the world on a flat surface can help establish their locations (both relative to each other and relative to the world).

*Binocular Cues:* Unlike 2D pictures, 3D virtual environments show images to both eyes, which allow stereoscopic cues to show depth. The visual system uses the slight differences in the two pictures (that are a result of the interpupillary distance) to gain significant information about spatial depth and location of objects in a scene.

*Object Shapes:* Another example of compensation for distorted objects by the visual system is when rigid and recognizable shapes are presented. Perkins [15] found that participants would compensate for oblique views of rectangular solids, and even when the judgments were inaccurate, they were only inaccurate by a small amount. If mental models are formed for commonly encountered objects, then it will be easier to match an image of such an object to the correct mental model, even if distorted.

## 4 Guidelines for Design

The main reason for examining the issues related to passive participants in a one-wall cave setting is the ability to better utilize these environments in groups. Traditionally, virtual reality systems have been solitary experiences. Complex hardware solutions have often allowed for two participants, but a group setting with a dozen users has been considered impractical other than in large theatrical settings or where the tracking is severely limited. We seek to exploit the idea of passive participants to create group VR applications and are exploring both the impact on passive participants and the transfer of active status among participants.

One major influence on design is a factor we are calling *leverage*. As seen in Equation 6, the amount of offset (skew) of the perceived point is scaled by the ratio  $z/z_a$ . As the distance from the screen increases, the offset also increases. As an example, if the active participant moves to the right 10cm, an object close in depth to the screen ( $z$  values near zero) will move very little. However, an object far from the screen, such as an object in the far distance, will move a very large amount.

Given this concept of leverage and the psychological concepts in Section 3, we are proposing several design guidelines for passive participant powerwall systems. These guidelines are intended to be used when distortions affect passive participation, to raise the overall utility of the system.

*Guideline 1:* Keep as much content at a depth near the screen as possible. This is the most obvious design guideline, since it minimizes the amount of leverage, and it keeps objects within interaction reach.

*Guideline 2:* When possible, the active participant should be close to the screen. An active participant close to the screen will minimize the ratio  $z/z_a$  in Equation 6 for user interface elements. Similarly, the active participant should remain closer to the center of the screen, so that the controlling viewpoint is the average of the possible passive participant viewpoints. This causes fewer overall discrepancies between viewpoints.

*Guideline 3:* Passive participants should be farther from the screen. Decreasing the angle of sight lines between the active and passive participant decreases the relative discrepancy in orientations. Moving users back limits the range of angles for these users, thus decreasing orientation confusion due to the differential rotation effect. In a setting where the active participation passes from user to user, this assignment may be based on physical position with a "move up to take control" type of approach.

*Guideline 4:* Avoid backgrounds or fix them in place. The problem with backgrounds is that they are naturally the farthest from the screen and, therefore, the most subject to leverage. Since they are often far in distance, users do not expect them to move that much and appear to be comfortable with this solution in preliminary experiments. This guideline also has the potential to minimize cybersickness effects that result from a moving background.

*Guideline 5:* Virtual objects should be familiar to users. Well-known or familiar objects appeal to long-established mental models. Because of this object recognition, users are better able to compensate for distortions in virtual objects. Overall, using familiar objects will help connect a virtual object to the user's sense of its physical presence in the real world. However, if novel objects are used, utilize Guideline 6.

*Guideline 6:* When familiar objects cannot be used to satisfy Guideline 5, emphasize pictorial cues such as shading, color fading, and atmospheric effects. This can help compensate for the lack of familiarity, allowing for novel virtual objects.

## **5 Future Work**

We are currently examining a new modality for virtual reality environments that is designed to optimize the experience for a group of users, rather than a single participant. Our first focus is discovering and alleviating the problems passive observers experience in immersive environments, such as perception conflicts for passive observers, as well as triggers for cybersickness. This work will then move beyond alleviating these problems, and leverage psychological and physiological compensatory mechanisms in order to deliver a better group interface for virtual environments. We are also exploring enhancements for a group virtual environment, such as efficient ways to transition between active and passive participants, creating a more satisfying collaborative, immersive, environment. To demonstrate the guidelines presented in this paper, we are designing an experiment that will demonstrate the ability of a user to sufficiently judge object characteristics in a powerwall system, despite being removed from the primary viewpoint.

## References

1. Ramsey, D.: 3D Virtual Reality Environment Developed at UC San Diego Helps Scientists Innovate (2009), <http://ucsdnews.ucsd.edu/newsrel/general/09-083DVirtualReality.asp> (accessed January 20, 2013)
2. Agrawala, M., Beers, A.C., McDowall, I., Fröhlich, B., Bolas, M., Hanrahan, P.: The Two-user Responsive Workbench: Support for Collaboration through Individual Views of a Shared Space. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1997), pp. 327–332. ACM Press/Addison-Wesley Publishing Co., New York (1997)
3. Fröhlich, B., Blach, R., Stefani, O., Hochstrate, J., Hoffmann, J., Klüger, K., Bues, M.: Implementing Multi-Viewer Stereo Displays. In: Proceedings: WSCG (Full Papers), pp. 139–146 (2005)
4. Kulik, A., Kunert, A., Beck, S., Reichel, R., Blach, R., Zink, A., Fröhlich, B.: Clx6: A Stereoscopic Six-User Display for Co-located Collaboration in Shared Virtual Environments. In: Proceedings of the 2011 SIGGRAPH Asia Conference, SA 2011 (2011)
5. Goldstein, E.B.: Perceived Orientation, Spatial Layout, and the Geometry of Pictures. In: Ellis, S.R., Kaiser, M.K., Gunwald, A.J. (eds.) Pictorial Communication in Virtual and Real Environments, ch. 31, pp. 480–485. Taylor & Francis Inc., Bristol (1991)
6. Goldstein, E.B.: Spatial Layout, Orientation Relative to the Observer, and Perceived Projection in Pictures Viewed at an Angle. *Journal of Experimental Psychology: Human Perception Performance* 13, 256–266 (1987)
7. Hancock, M., Nacenta, M., Gutwin, C., Carpendale, S.: The Effects of Changing Projection Geometry on the Interpretation of 3D Orientation on Tabletops. In: ACM International Conference on Interactive Tabletops and Surfaces, New York, pp. 157–164 (2009)
8. Chen, Y., Dong, X., Hagstrom, J., Stoffregen, T.A.: Control of a Virtual Ambulation Influences Body Movement and Motion Sickness. In: *BIO Web of Conferences*, vol. 1(16) (2001)
9. Kim, Y.Y., Kim, H.J., Kim, E.N., Ko, H.D., Kim, H.: Characteristic Changes in the Psychological Components of Cybersickness. *Psychophysiology* 42(5), 616–625 (2005)
10. Dong, X., Stoffregen, T.A.: Postural Activity and Motion Sickness Among Drivers and Passengers in a Console Video Game. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54(18), 1340–1344 (2010)
11. Christou, C., Parker, A.: Visual Realism and Virtual Reality: A Psychological Perspective. In: Carr, K., England, R. (eds.) *Simulated and Virtual Realities*, ch. 3, pp. 53–84. Taylor & Francis (1995)
12. Cognitive Systems. *Spatial Cognition* (2012), <http://www.spatial-cognition.de> (accessed January 19, 2013)
13. Marr, D.C.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, New York (1983)
14. Vishwanath, D., Girshick, A.R., Banks, M.S.: Why Pictures Look Right When Viewed from the Wrong Place. *Nature Neuroscience* 8(10), 1401–1410 (2005)
15. Perkins, D.N.: Compensating for distortion in viewing pictures obliquely. *Attention, Perception, & Psychophysics* 14(1), 13–18 (1973)

# Virtual Reality Based Interactive Conceptual Simulations

## Combining Post-processing and Linear Static Simulations

Holger Graf and André Stork

Fraunhofer Institute for Computer Graphics, Fraunhoferstr. 5, 64283 Darmstadt, Germany  
{holger.graf, andre.stork}@igd.fraunhofer.de

**Abstract.** This paper presents a new approach for the design and realization of a Virtual Reality (VR) based engineering front end that enables engineers to combine post processing tasks and finite element methods for linear static analyses at interactive rates. “What-if-scenarios” have become a widespread methodology in the CAE domain. Here, designers and engineers interact with the virtual mock-up, change boundary conditions (BC), variate geometry or BCs and simulate and analyze its impact on the CAE mock-up. The potential of VR for post-processing engineering data enlightened ideas to deploy it for interactive investigations at conceptual stage. While it is a valid hypothesis, still many challenges and problems remain due to the nature of the “change’n play” paradigm imposed by conceptual simulations as well as the non-availability of accurate, interactive FEM procedures. Interactive conceptual simulations (ICS) require new FEM approaches in order to expose the benefit of VR based front ends.

**Keywords:** Computer Aided Engineering, Interactive Conceptual Simulations, VR environments for engineering.

## 1 Introduction

“What-if-scenarios” (conceptual simulations) have become a widespread methodology within the computer aided engineering (CAE) domain. Here, designers and engineers interact with the virtual mock-up, change boundary conditions (BC), variate geometry or BCs and simulate and analyze its impact on the CAE mock-up. The potential of VR for post-processing engineering data enlightened ideas to deploy it for interactive investigations at conceptual stage (interactive conceptual simulations - ICS). It is still a valid hypothesis, while many challenges and problems remain. The conceptual stage during a design is inherently driven by the nature of the “change’n play” paradigm. Coupling these with Finite Element Methods (FEM) imply new solutions and optimizations in view of the current non-availability of accurate, interactive FEM procedures for interactive processing. VR is predominately used for data visualization of scientific raw data. Therefore, classical solutions still use scientific visualization techniques for large data visualization [1] or use interpolated pre-computed result data sets, e.g. [2,3] for interactive investigations. Both approaches imply a bottleneck of

data set processing, filtering and mapping and impose restrictions to the processing capability of the underlying system thus influence the turn-around loop of simulation and visualization. Thus, conducting a CAE analysis, steered from a VR environment is a different story and only few research work exist, e.g. [4,5,6]. Classical CG methods are too limited due to the simplified, underlying mathematical models for real-time analysis [7]. In fact several approaches are driven by visual appearance rather than physical accuracy needed within engineering environments. Major attention has been given to the area of *deformable object simulations* in the past [8]. Here, the challenge is to solve the underlying system of differential equations imposed by the physical phenomena modeled by Newton's second law of motion. The approaches make typically use of explicit time integration schemes, fast in evaluation and small in computational overhead, e.g. [9]. Implicit time integration schemes which usually lead to a more stable calculation of the results for solid deformation are based on complete assemblies into large systems of algebraic equations, which might be solved using pre-processing techniques (such as matrix pre-inversion) [10,11], or the conjugate gradient method eliminating corotational artifacts, e.g. [12,13]. A combination of several "best" practices for physical simulations has been published recently in [14] with a dedicated focus on fast solutions being robust to inconsistent input, amenable to parallelization, and capable of producing the dynamic visual appearance of deformation and fracture in gaming environments. However, the scope of all mentioned methods cannot handle more than very few thousand elements or are too imprecise for engineering analyses. So the main question to be answered remains: how can interactive FEM methods be designed and realized at conceptual stage that are efficient with respect to time consumption, computer resources and algorithmic complexity but at the same time result in an accurate and robust simulation?

## 2 Concept

Aim of our approach is to provide the possibility to couple post processing tasks with a simulation engine, that allows for any interaction performed by the end user to update the simulation results in real-time at the same time to perform an analysis. Here, we are focussing on a direct link of typical post-processing metaphors such as cross sectioning which should provide an insight into the object while simulating the model.

Typically the user wants to move a load case from one node position to another one yet being unrestricted to the number of nodes within the load case. For cross sectioning the plane that cuts through the object will be orthogonal to the device of the end user: In our case a flying pen. Moving the device results in an update of the position of the cross section. A re-simulation of the mock-up should then be performed instantly. However, the use of VR environments implies an intervention with the simulation engine at update rates for (re-)simulation-visualization loops at 20-30 fps, i.e. 0.05 secs. This in turn requires direct access to the underlying mathematical procedures respectively finite element methods.



We headed for a concept based on a classical CAE/VR process chain using a distributed software architecture [15]. Typically, we hold model presentation in a VR client as a surface model being pre-processed by scientific visualization techniques (i.e. extracting the outer domain for visualization, filtering and mapping of results to color scales, etc.). The overall volumetric CAE mock-up is kept on a dedicated simulation service that accounts for linear static analysis.

Thus, the coupling of post processing tasks with the simulation engine requires operations/interactions being performed in the VR client (e.g. moving a load vector/user force) being mirrored to the simulation services. Ideally the simulation engine might be based on optimized FE-methods that could comply with the requirements of the post-processing tasks. I.e., if an engineer uses cross sectioning through the model, the simulation engine would only need to calculate for the “visible” elements (this means the element in the current view frustum). This leads conceptually to a reduction of the solution space, thus a reduction of the system of linear equations that needs to be solved. Another challenge will be imposed by aiming at changing geometrical features in the mock-up (e.g. through holes). Moving features provide an insight to Any changes done at surface level need to be reflected within the volumetric mock-up in the simulation service.

### 3 Realisation

#### 3.1 ICS at Boundary Condition Level

The realisation for conceptual simulations at BC level is based on a methodology introduced in earlier work using a pre-processing step [11]. This method uses an inversion of the underlying stiffness matrix  $A$  via a *preconditioned minimal residual method* for the overall linear static equation  $\mathbf{u} = \mathbf{A}^{-1} \cdot \mathbf{I}$ , with  $\mathbf{u}$  being displacement and  $\mathbf{I}$  load vector.

The iterative scheme is given by minimizing a Frobenius norm, i.e. it minimizes the functional  $F(B) = \|\mathbf{I} - \mathbf{B}\mathbf{A}\|_F^2$  with  $B$  being the sought inverse to  $A$ . Splitting the functional into components, the scheme seeks a solution  $\hat{f}_j(\underline{b}_j) = \min_j \|\underline{e}_j - \mathbf{A}\underline{b}_j\|_2^2$ ,  $j=1, \dots, n$  deploying a CG-iterative method [16]. Completing the inversion of the matrix for a given error threshold, a dedicated stop criterion provides the envisaged precision. This can be adjusted by the engineer himself. Due to the nature of iterative schemes, the precision significantly influences the computation time of the matrix inversion, thus the availability of the model to be inspected. Therefore, this step is done within an offline preparation mode. However, once the model is available, the engineer has several degrees of freedom to investigate conceptual changes at boundary condition level. Here, we have managed to reduce the solution to a simple matrix-vector multiplication.

During the real-time calculation step within the VR client by interactively moving the load case,  $\mathbf{I}$  is dynamically filled with values according to force and direction by

the position/orientation of the user’s interaction device (fig. 2). The simulation filters all unnecessary elements and related rows in the matrix ( $a_{ij} := 0, \forall j \in \{s | l_s = 0\}$ ) (marked black – fig. 1, left), thus, takes only those elements into account that contribute to the results. A second optimisation uses a view dependent element masking technique by neglecting the affected rows within the inverted matrix. As a consequence, a further acceleration and turn-around loop speed-up of the matrix-vector multiplication is feasible. In order to include only those elements visible to the user into the computation, an additional occlusion evaluation step as to which elements are within the view direction of the viewer and which are occluded has to be performed (marked grey – fig. 1, right).

$$\begin{array}{c}
 \mathbf{u} = \mathbf{A}^{-1} \mathbf{X} \\
 \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_8 \end{bmatrix} = \begin{bmatrix} * & \blacksquare & * & \blacksquare & \blacksquare & \blacksquare & * \\ * & \blacksquare & * & \blacksquare & \blacksquare & \blacksquare & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \blacksquare & * & \blacksquare & \blacksquare & \blacksquare & * \end{bmatrix} \begin{bmatrix} l_1 \\ 0 \\ l_3 \\ l_4 \\ \vdots \\ 0 \\ 0 \\ l_7 \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_6 \\ u_7 \\ u_8 \end{bmatrix} = \begin{bmatrix} * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \\ * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \\ * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \\ * & \blacksquare & * & * & \blacksquare & \blacksquare & * & \blacksquare \end{bmatrix} \begin{bmatrix} l_1 \\ 0 \\ l_3 \\ l_4 \\ \vdots \\ 0 \\ 0 \\ l_7 \\ 0 \end{bmatrix}
 \end{array}$$

**Fig. 1.** The used simulation scheme for solving the system of linear equations: throwing away useless values and reducing the matrix-vector calculation load (left); Results of element masking using an occlusion evaluation and taking into account only visible elements (right) [11]

The major advantage of the viewdependent masking due to an occlusion evaluation results in a direct exploitation for post processing tasks.

### 3.2 Implementation

With respect to the UI concept of a direct interaction method within the *VR client* ( $\text{VSC}::\text{VR}::\text{SG} := \text{pVRservice}^1$ ) simple user interactions based on selection boxes provide a mechanism to assign loads on a surface or a group of elements. The system supports different kinds of loads (BCs) that can be attached or might even be deleted to/from nodes. In order to reflect the changes of the VR client within the instance of the *simulation service* ( $\text{VSC}::\text{RT} := \text{pCAEService}$ ), the methods can be accessed by asynchronous calls to the remote service instance in order not to block the current visualization process. Adequate methods are provided by each service instance through their interface.

<sup>1</sup> Within the implementation the different services are represented by a service instance ( $\text{p}\{X\}\text{Service}$ ,  $X \in \{\text{VR}; \text{CAE}\}$ ) providing adequate interfaces to other services of the distributed system.  $\text{VSC}::\text{VR}::\text{SG}$  itself is the client service,  $\text{pVRService}$  an instance of it.

As mentioned above, the user interactions performed on the mock-up are not restricted to a pure visualisation of the results. The user is able during post processing tasks to define cross sections in which the user is able to newly mask the volume taking into account only the visible elements on the surface. Therefore, the `pCAEService` instance allows extracting the outer surface of the mock-up taking into account the position of the pen and the plane being orthogonal to the pen (`pPosition`, `pNormal`). Having fixed the cross section position, the user is able to trigger instantly a re-simulation through a change of magnitude and direction of the user force vector (`pUser_Force_Vector`), allowing a view insight the volume's newly simulated stress field. The magnitude of the user defined force can still be varying.

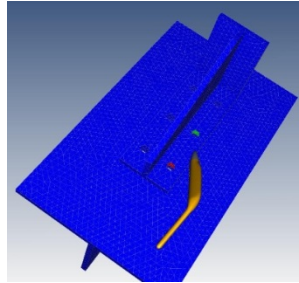
```
// Operations during the second operational phase (online simulation)
// Within the VR client capture position, orientation and normal of the
// interaction device during post processing as well as the magnitude of
// device movement and establish pUser_Force_Vector
// Extract the cutting plane and mask the resulting elements (BC_MASK)
// for updating the simulation and visualisation (updateRTCalculation)

while(pPosition, pNormal) //- moving the interaction device
{
    pCuttingPlane = pVRService -> updateCrossSection
        (pCAEService, pFlag, pPosition, pNormal)
    pVRService -> updateElementMasks(pCAEService, ELEMENTS, pCutting
        Plane->getNodes());
    pVRService -> updateElementMasks(pCAEService, BC_MASK pCutting
        Plane->getNodes());
    pMesh = pVRService -> updateRTCalculation(pCAEService, pFlag,
        pUser_Force_Vector);
    pVrService -> visualise(pMesh);
}
```

### 3.3 ICS at Geometrical Level

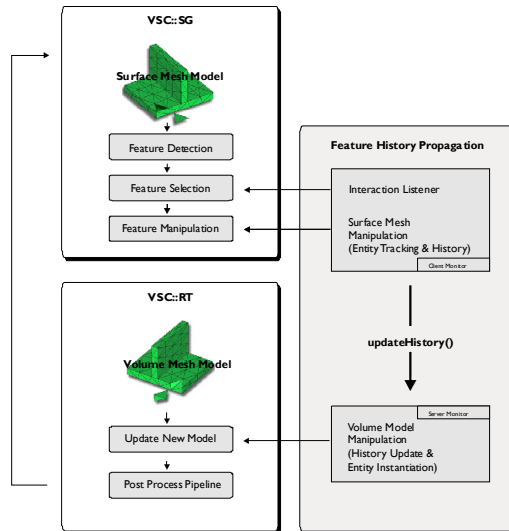
This section covers interactive modifications of the engineering domain, i.e. geometrical respectively topological level. The implemented techniques for mesh manipulation of the surface/volume mesh are classified as *feature dragging* or surface/volume re-meshing techniques. The process consists in moving vertices of selected simplicial elements causing mesh collapse/split operations in a way that allows the consistency of the mesh being kept using local topological operations. Those operations deal with adapting ill-shaped areas of the given simplicial mesh to a well-shaped area of it.

Of course adequate metrics for quality evaluation of the affected elements during a user defined movement of a group of vertices have to be used in order to perform topological operations for those elements with a low quality (i.e., a quality lower than a given threshold), e.g. [17]. The ill-shaped simplexes are then modified by operations such as edge collapse, edge split, tetrahedron collapse-face swap which are applied appropriately, depending on the damage or degeneracy of the elements in question. The smallest and biggest edges (as well as sliver tetrahedron, i.e. flat tetrahedrons) are adapted to maintain a mesh with a good quality and preserving the consistency for the newly triggered simulation run. This is necessary due to the fact that slight changes of the underlying mesh and thus the associated interpolation schemes might have significant impact on the results and the quality of the simulation.



**Fig. 2.** Interactive feature dragging of through holes (according to pen movement)

In order to be able to move selected features within a mesh, an update mechanism of the manipulations performed on VR client side and its propagation back into CAE service has to be established. As our VR environment processes surface meshes, e.g. the outer surfaces of a CAE mock-up, we need to be consistent with the different object instantiations in the system, several interactions and manipulations done within the client have to be propagated to the `pVRService` and/or `pCAEService`. The methodology used for the conceptual simulation process with respect to feature movement is shown in fig 3. This mechanism allows the feature movement being extended for real conceptual simulations. The feature history propagation is divided into different routines which are synchronised using the asynchronous method `updateHistory()`.



**Fig. 3.** Feature synchronisation between VR client (`pVRService`) and simulation service (`pCAEService`)

On client side a monitoring mechanism, collecting the modifications done on entity level controls the surface mesh manipulations and records user interactions, i.e. change of entities, displacements of nodes, edges, connectivity information, etc. On service side the monitoring mechanism is responsible for a volume mesh instantiation of the performed actions recorded on client side. As the instantiation does not need to be performed in real-time, it is done after the interaction terminates triggering the `updateHistory()` method.

### 3.4 Implementation

In general, a typical dragging operation is performed based on a change of selected vertice positions belonging to a feature. The displacements of those vertices entering a new position within the higher level compound of elements (face) trigger a remeshing according to the decisions taken by the quality measure. Several vertices can be classified and marked as those belonging to a feature (“SELECTED”) and those belonging to a fixed part of an area (“FIXED”). As the positions of feature vertices change during dragging operations the vertices undergo a penalty criteria as to which a further collapse, swap or insert operation will be conducted. A special area of interest around the feature is the one containing vertices or edges of the feature as well as vertices or edges that belong to a fixed part of the compound face. Therefore, a further flag for element vertices and edges as “SHARED” indicate that several dragging operations are only performed on those (“SHARED”, “SELECTED”) and lead to the envisaged remeshing. After a principle topological modification of elements their vertices are marked as “KILLED”, “SELECTED”, “SHARED” or “FIXED” depending on whether vertices are deleted (i.e. during a collapse operation) or further used for operation (i.e. during a split process). The realization sequence in pseudo code looks like:

```
// Selection of entities
    defineFeatureBoundary(in pPosition, in pOrientation, in
        p{element_heuristics}, out Q p{entitiy_selection});

// label the boundary elements
    labelBoundary(in Q p{entitiy_selection}, out Q p{marked_entities});

// inquire for "SHARED" and "SELECTED" elements and define buffer of
// elements that listen to modifications triggered by VSC::EVT events
    getSharedBoundaries(in Q p{marked_entities}, out
        Q p{buffer_entities});

// translate pPosition and pOrientation of the pen into displacements of
// the vertices for the buffer_entities and record movements

while(pPosition, pOrientation) // -- Start movement of the pen
{
    calculateMovement(in pPosition, in pOrientation, out pDirection,
        out pStart, out pDistance);

// Topological operations are applied to every boundary resp. buffer
// vertex.
    applyMovementsToSurface(in Q p{buffer_entities}, in pDirection ,
        in pDistance)
```

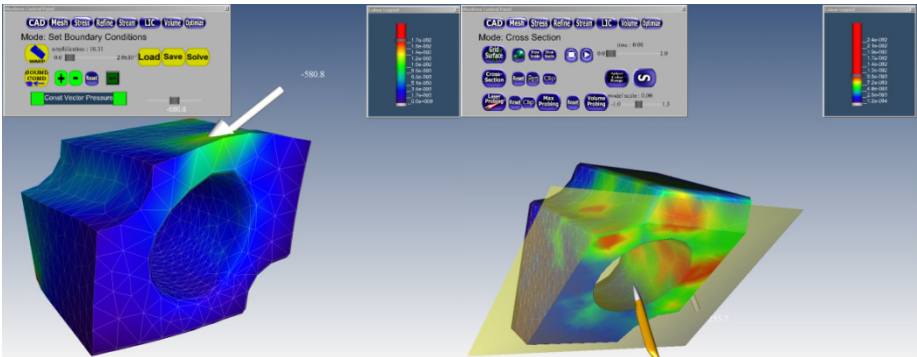
```

{
  ∀ pEntities ∈ p{buffer_entities}
  {
    // edge_collapse, edge_swap, edge_split process
    adaptiveRemesh(in pEntities, in pDirection, in
                  pDistance);
    labelEntities(in pEntities, out Q
                 p{new_buffer_entities});
  }
}
// Update the remote service with modified entities and instantiate
// changes on the surface into the volume mesh
pVRService::SG -> sendc_updateHistory(pCAEService, pDirection,
                                     pDistance, pStart, Q p{new_buffer_entities});

```

## 4 Results

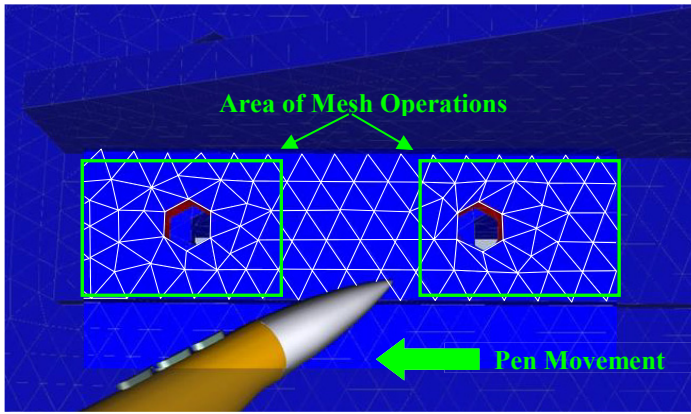
The presented system enables us to link typical post-processing tasks (e.g. cross sectioning, etc.) directly to the simulation engine evaluating the viewpoint and current force vector of the engineer. He is then able to define cross sections enabling to newly mask the volume taking into account only the visible elements on the surface (see fig. 4). A re-simulation due to a change of the magnitude and direction of the user force vector allows a view insight the volume's stress field. The magnitude of the user defined force can still be varying.



**Fig. 4.** Integrated post-processing and simulation; left: resulting deformations (scalable post-processing of displacement field); right: results of a cross-section simulation with an update of element masks

The integration of a conceptual change of a feature position within a given design domain into the framework follows the mechanism described in above using the feature history propagation. As a result, it enables the user to mark and select certain features in the domain on client side and “drag” them in 3D space to another position. As the manipulations within the `VSC::SG` client are performed on the surface

representation, the volume mesh kept in the `VSC::RT` service backbone has to be updated accordingly. The selection of mesh entities is based on face identification. Several compound elements belonging to a compound of faces with heuristically similar characteristics or a CAD face can be selected (see fig. 5). During a spatial change of selected features, i.e. through holes, on client side (`VSC::SG`) a monitoring mechanism collecting the modifications done at entity level controls the surface mesh manipulations and records user interactions, i.e. change of entities, displacements of nodes, edges, connectivity information, etc. They are then propagated back to the `VSC::RT` service.



**Fig. 5.** Interactive feature dragging: two through holes (marked as red) and resulting mesh operations at surface level being propagated to the volume simulation engine

## 5 Conclusion

This paper presents a new approach for the design and realization of a Virtual Reality (VR) based engineering front end that enables engineers to combine post processing tasks and finite element methods for linear static analyses at interactive rates. The engineer is able to steer post-processing analysis and re-simulation “at his fingertip”. The implementation has been done within a distributed set-up in order to comply with the limitations of CAE simulations and their mock-ups. Several operations can be performed in real-time for selected domains. However, the model size cannot be arbitrary as shown in [11]. This might impose a critical limitation to the use of the system for larger models. We therefore are working on subdomaining mechanism that might reduce the overall domain in order to provide also a certain scalability of the system. Yet, we are optimistic that the presented ICS might enable engineers to use this paradigm for “what-if-analysis” in order to be capable of answering the question: “*where do I have to spend my analysis time?*”. As further future work we head towards a closer interlink between mesh and simulation. Thus, exploiting the neighborhood relationships between nodes might lead to an optimization of the stiffness matrix entries that might eventually lead to a significant reduction in computational time.

## References

1. Scherer, S., Wabner, M.: Advanced visualization for finite elements analysis in virtual reality environments. *International Journal on Interactive Design and Manufacturing* 2, 169–173 (2008)
2. Stork, A., Thole, C.A., Klimenko, S., Nikitin, I., Nikitina, L., Astakhov, Y.: Simulated Reality in Automotive Design. In: *Proc. of IEEE International Conference on Cyberworlds 2007*, pp. 23–27 (2007)
3. Lee, E.J., El-Tawill, S.: FEMVrml: FEMvrm: An Interactive Virtual Environment for Visualization of Finite Element Simulation Results. *Advances in Engineering Software* 39, 737–742 (2008)
4. Vance, J.M., Ryken, M.J.: Applying virtual reality techniques to the interactive stress analysis of a tractor lift arm. *Finite Element Analysis and its Design* 35(2), 141–155 (2000)
5. Connell, M., Tullberg, O.: A Framework for the Interactive Investigation of Finite Element Simulations within Virtual Environments. In: *Proc. of the 2nd International Conference on Engineering Computational Technology*, Leuven, Belgium (2000)
6. Connell, M., Tullberg, O.: A Framework for Immersive FEM Simulations using Transparent Object Communication in Distributed Network Environments. *Advances in Engineering Software* 33(7-10), 453–459 (2002)
7. Georgii, J.: Real-Time Simulation and Visualisation of Deformable Objects. Dissertation, Institut für Informatik, Technische Universität München (TUM) (2008)
8. Nealen, A., Mueller, M., Keiser, R., Boxerman, E., Carlson, M.: Physically Based Deformable Models in Computer Graphics. *Computer Graphics Forum* 25(4), 809–836 (2006)
9. DeBunne, G., Desbrun, M., Barr, A., Cani, M.-P.: Dynamic real time deformations using space & time adaptive sampling. In: *Proceedings of SIGGRAPH 2001*, pp. 31–36 (2001)
10. Bro-Nielsen, M., Cotin, S.: Real-time volumetric deformable models for surgery simulation using finite elements and condensation. In: *Proceedings of Eurographics 1996*, pp. 57–66 (1996)
11. Graf, H., Stork, A.: Linear static real-time finite element computations based on element masks. In: *Proc. of the ASME 2011 World Conference on Innovative Virtual Reality, WINVR 2011*, Milan, Italy (2011)
12. Müller, M., Dorsey, J., McMillan, L., Jagnow, R., Cutler, B.: Stable real-time deformations. In: *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 49–54 (2002)
13. Müller, M., Gross, M.: Interactive virtual materials. In: *Proceedings of Graphics Interface*, pp. 239–246 (2004)
14. Parker, E.G., O'Brien, J.F.: Real-time deformation and fracture in a game environment. In: *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, New Orleans (2009)
15. Graf, H.: A "Change'n Play" Software Architecture Integrating CAD, CAE and Immersive Real-Time Environments. In: *Proc. of the 12th International Conference on CAD/Graphics*, Jinan, China (2011)
16. Chow, E., Saad, Y.: Approximate Inverse Preconditioners via Sparse-sparse Iterations. *SIAM J. Sci. Comput.* 19(3), 995–1023 (1998)
17. Shewchuk, J.: What is a Good Linear Element? Interpolation, Conditioning and Quality Measures. In: *Proc. of the 11th International Meshing Roundtable*, pp. 115–126 (2002)



# Enhancing Metric Perception with RGB-D Camera

Daiki Handa, Hirotake Ishii, and Hiroshi Shimoda

Kyoto University, Graduate School of Energy Science, Kyoto pref., Japan  
{handa,hirotake,shimoda}@ei.energy.kyoto-u.ac.jp

**Abstract.** Metric measurement of environment has fundamental role in tasks such as interior design and plant maintenance. Conventional methods for these tasks suffer from high development cost or unstability. We propose a mobile metric perception enhancement system which focuses on interactivity through user locomotion. The proposed system overlays geometric annotations in real-time on a tablet device. The annotation is generated from RGB-D camera in per-frame basis, alleviating the object recognition problem by effectively utilizing processing power of human. We show a few illustrative cases where the system is tested, and discuss correctness of annotations.

**Keywords:** Augmented Reality, Augmented Human, Mobile Device, RGB-D Camera, Geometric Annotation, Per-frame Processing.

## 1 Introduction

Real world tasks such as interior design and plant maintenance rely on knowledge of geometric properties of surrounding objects. In these scenarios, measurement of environment often forms the basis of higher level sub-tasks. We propose metric perception enhancement through overlaying geometric annotation extracted from RGB-D data in real-time.

Existing augmented reality solutions for these tasks mostly depend on the idea of overlaying suitable pre-made virtual objects such as furniture or CAD model [1]. While this approach can potentially provide tailored user experience, these applications tend to add little benefit compared to required application development and deployment cost. These costs may occur from employment of artists to create virtual object, setup of markers to track the device, or maintainance of up-to-date CAD data of the environment.

On the other hand, there are many methods to create 3D model of the environment on-the-fly, generally called Simultaneous Localization and Mapping (SLAM) [2]. But these methods are either not robust enough, only able to provide sparse model, or computationally expensive. So dynamic content creation through automatic modeling of environment is not feasible.

Recent introduction of consumer-grade RGB-D sensors such as Microsoft kinect enable us to use it on mobile devices such as tablets, making it possible to robustly acquire local 3D point cloud in real-time.

We propose geometric annotation application that can be used with little constraint on the environment. We generate salient annotations from RGB-D data, and overlay them on per-frame basis. The proposed system can annotate straight line in sight with lengths, and surfaces with contour lines. While the quality of output is lower than that of perfect CAD data, tight interaction loop created by per-frame presentation of data can compensate the downside, and can provide reasonably good user experience at very low cost. The key insight is that human can easily associate flickering or duplicated annotations to real world structure, while it is very difficult for computers to accurately create coherent model of the environment from raw data.

We will illustrate a few cases where our system would be useful and discuss correctness of generated annotations.

## 2 Generating Annotation

To visualize metric properties of environment, two complementary kinds of annotations are generated. The *length annotations* enable the user to perceive lengths of straight edges abundant in artificial objects. The *height annotations* are contour lines to help understanding featureless or curved surfaces where straight edges are absent and thus length annotation is unavailable.

Dataflow of the annotation process is shown in Fig. 1. The input of the system is QVGA frames from a RGB-D camera and the gravity vector<sup>1</sup> from a tablet.

### 2.1 Edge Detection and Refinement

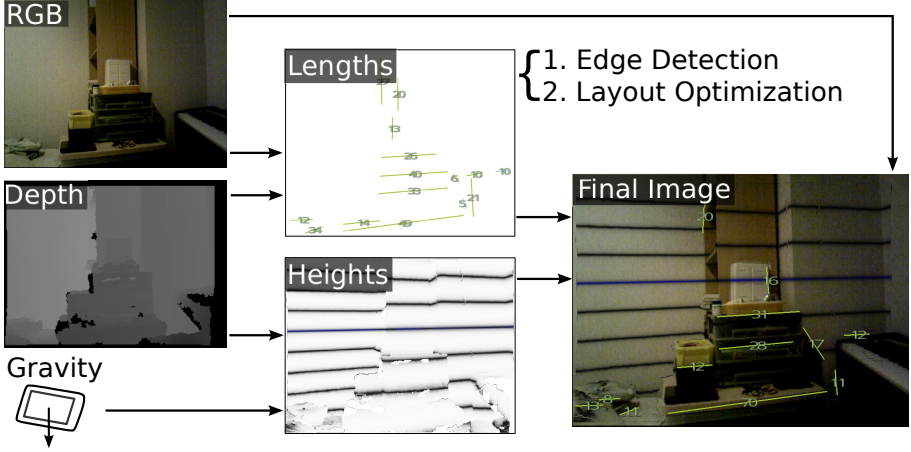
The goals here are extraction of straight edges and calculation of their lengths. Output of current depth sensors typically contains jaggiens of several pixels near object outline, while RGB image have effective angular resolution of nearly one pixel. So, three-dimensional edges are estimated from line segments in RGB image.

Line segments are extracted from RGB image by first converting it to grayscale, and then applying LSD [3] detector. The detected line segments contain *Number of False Alarms* (NFA) values, which are used as saliency in later optimization phase.

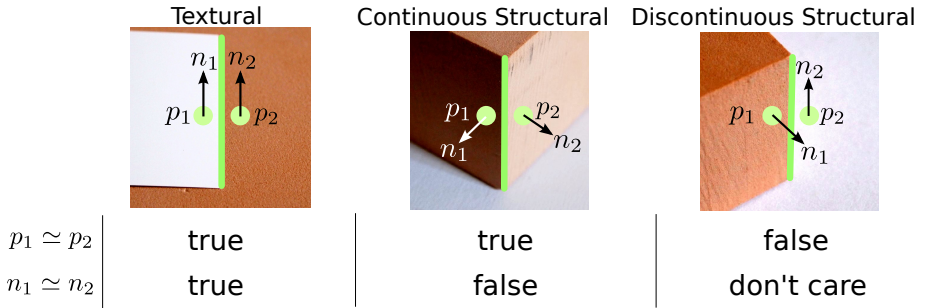
Detected edges can be categorized to three classes as shown in Fig. 2; a textural edge lies on planar surface, and a structural edge corresponds a ridge or a cliff of an object. Structural edges are further divided to continuous or discontinuous by whether two sides of the edge are on a same object (continuous) or not (discontinuous). Discontinuous structural edges need special treatment when calculating length, since depth is ill-defined on the discontinuous edge.

---

<sup>1</sup> Mobile platform such as Android provides gravity sensor based on low-pass filtering of accelerometer data.



**Fig. 1.** Upper Middle: Lengths from line segments, Lower Middle: contours from per-pixel depth coloring



**Fig. 2.** Edges can be classified by comparing positions and normals near midpoints. In reality, occlusions, shadows and noise make distinction unclear.

To check discontinuity of an edge,  $p_1 \simeq p_2$  condition (in Fig. 2) is used. When depth is continuous at an edge, 3D distance  $d$  between two symmetric points near the midpoint is linear to that of screen space. Pair-distance  $d(s)$  for points  $2s$  apart in screen space is defined as follows:

$$d(s) = |T(p_{\text{mid}} + sn) - T(p_{\text{mid}} - sn)| \quad (1)$$

where  $p_{\text{mid}}$  is the midpoint of the edge,  $T(p)$  is 3D position of the pixel, and  $n$  is the normal of the segment. By using  $d(s)$ , the discontinuity condition is approximated by  $\frac{d(5\text{px})}{d(2\text{px})} < \frac{5}{2}\alpha$ , where  $\alpha \simeq 1$  is a sensitivity constant.

After edge classification, discontinuous edges are refined by moving toward the nearer (i.e. foreground) side to avoid jagged region. After edge refinement,

length is calculated respectively from two endpoints of the segments. If depth at an endpoints is unavailable due to depth camera limitation, the edge is discarded as false one.

## 2.2 Layout Optimization

In complex scenes, edge annotations may become unreadable due to overlap. To mitigate this problem, annotation density distribution on screen is represented by a lattice, and edges are picked sequentially in order of decreasing saliency. The greedy selection process is depicted in the following pseudocode:

```
def select_edges(edges):
    bool[] [] density = {{false,...},...}
    edges_to_show = []
    for edge in sort(edges, order_by=NFA, decreasing):
        if not any(density[x,y] for (x,y) in cells_on(edge)):
            for (x,y) in cells_on(edge):
                density[x,y] = true
            edges_to_show.add(edge)
    return edges_to_show
```

Here we use 20 px for cell and lattice size where frame size is 320 px  $\times$  240 px. To allow edges with a shared vertex like a corner of a box, cells corresponding to endpoints are excluded when computing `cells_on(edge)`.

## 2.3 Height Annotation

Normalized gravity vector  $n_{\text{gravity}}$  is used to show contour lines. To draw a single contour line with camera-relative height  $h$ , intensity  $I(p)$  at pixel  $p$  in screen coordinates is determined by Eq. 2.

$$I_h(p) = \frac{1}{1 + (T(p) \cdot n_{\text{gravity}} - h)^2 w^2} \quad (2)$$

where  $T(p)$  is 3D position of  $p$  in camera coordinates, and  $w$  is a constant controlling the line width. In this paper, height annotations are drawn with 20 cm interval.

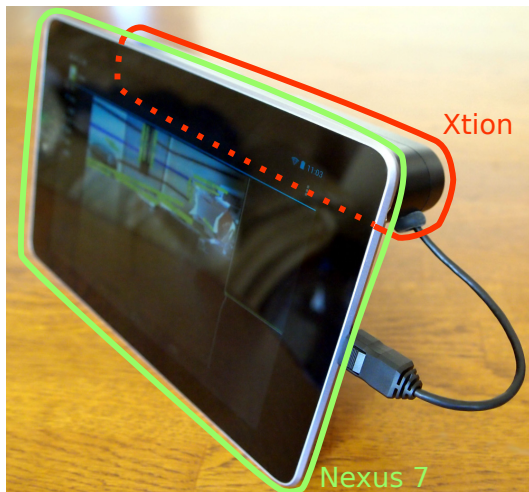
The decision to draw height annotations relative to device position instead of automatically detected floor, ensures smooth temporal behavior of lines by avoiding non-robust floor detection step. It is up to the user to hold the device at appropriate height to get meaningful readings.

## 3 Implementation

In this section, implementation details which can affect performance and mobility are described.

### 3.1 Hardware

The device consists of an Android tablet and a RGB-D camera as shown in Fig. 3. Since the camera is powered by USB from the tablet, there is no need for an external power supply. Mobility of the system is further increased by modifying the camera shell and cable. This results in a device with total weight of under 450 g, which can be used portably with a single hand.



**Fig. 3.** Nexus 7 tablet and modified ASUS Xtion PRO LIVE RGB-D camera connected via USB

Note that a Nexus 7 contains an accelerometer, so the only external component is the RGB-D camera.

### 3.2 Software

The system is implemented on Android 4.2.1, and most part is coded in Java. A screenshot in Fig. 4 shows the UI and a typical result of annotation.

To maximize performance, the line segment detector [3] is compiled for ARM NEON instructions and called via Java Native Interface. Rendering of annotations is performed on GPU, and particularly, height annotation is implemented as a fragment shader.

The UI allows respective switching of length and height annotations to increase framerate by turning off unnecessary annotations. To limit the mode of interaction to moving in the real world, controls for parameters such as detection threshold are intentionally excluded.

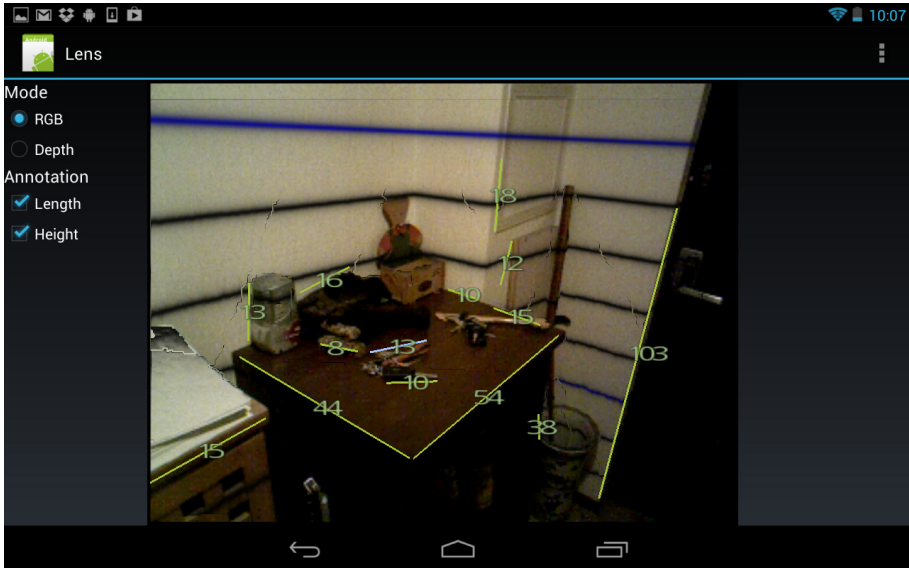


Fig. 4. A screenshot of the system

## 4 Experimental Evaluation

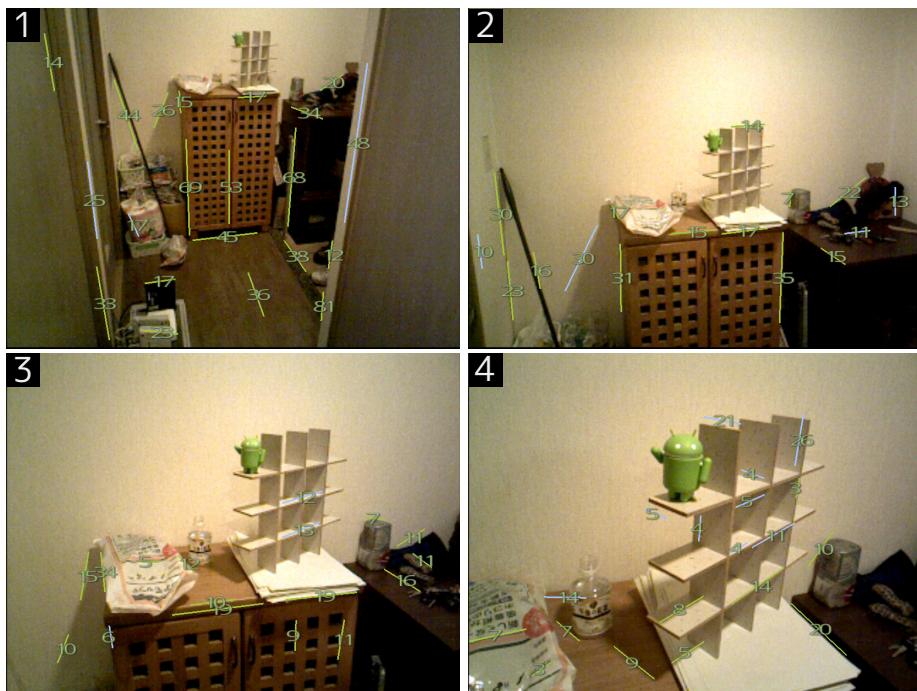
We illustrate several use cases by showing example of operation and evaluate accuracy of annotations. All examples were run at real-time frame rate.

### 4.1 Interactive Usage

Figure 5 shows the change in display when the user moved toward an object. Initially invisible small features (e.g. lattice-like object in 4) become visible with a closer look. In this example, natural user movement cause scale to change and show what the user would want to see. In general, it is often possible to read the length of an arbitrary edge by viewing from an appropriate angle and position. It can be argued that this kind of minimal-guessing (on the computer side) approach is more effective and feasible than trying to acquire detailed model of the environment and construct a GUI to choose what to see in the model.

Figure 6 illustrates how height annotation can complement length annotation for a curved object.

In these cases, two kind of annotations are used separately to see the effect respectively. Using both annotations simultaneously as in Fig. 4 does not cause a clutter, so we can omit GUI switches and make real world locomotion a sole, yet complete mode of interaction. This property would be useful when using the proposed technique with a HMD or a mobile projector like [4].



**Fig. 5.** 1-4: Scale of length annotation changes as the user moves toward the Android mascot

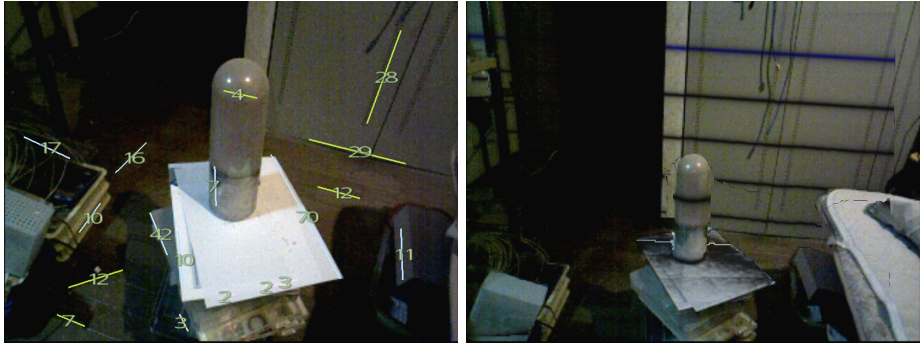
## 4.2 Latency

Important to interactivity is the latency. Typical latency to process a single frame is shown in table 1. Note that actual framerate is somewhat higher than determined by the total latency, since the code is multi-threaded.

**Table 1.** Typical Latency

Section	Time[ms]
Line Segment Detection (QVGA)	481
Edge Analysis & Refinement	8
Layout Optimization	4
Rendering & CPU-GPU Transfer	25
Total	518

Line segment detection is taking significant time and clearly needs a faster implementation, possibly on GPU. However, the system runs at nearly 30 fps when only height annotation is used.

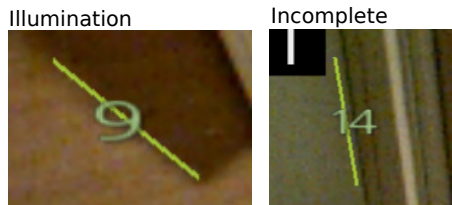


**Fig. 6.** Left: Length annotation cannot display height of the round-end cylinder Right: Height annotation reveals height of 1.5 units, which corresponds to 30 cm

### 4.3 Discussions on Correctness

Ultimately, precision would be limited by depth camera error, for which a detailed analysis exists [5]. However, incorrect lengths from false edges are far more noticeable in the current implementation.

In Fig. 5, there are roughly two kinds of false edges; edges corresponding to no structure nor texture, and fragmented or incomplete edges along long lines. An example for each kind is shown in Fig. 7.



**Fig. 7.** Left: false edge along shadow Right: edge is structural, but too short

The former is caused by shadow or gradation due to illumination, but human is so good at distinguishing illumination and texture (i.e. *lightness constancy* effect [6]) that difference between human and machine perception becomes noticeable. This kind of false edges are relatively harmless since they appear where other real edges are absent.

The latter is more problematic, since shorter edges can hide original long edge in layout optimization. Solution to this would be giving long edges higher scores in optimization, or using depth-guided line segment detection.

## 5 Conclusion

In this paper, we have shown that the conveying geometric information directly to the user is useful in various settings and relatively simple to implement com-



pared to conventional approaches like in [1]. We proposed a method to annotate lengths and contours and implemented in a truly mobile way.

The experiment shows the ability of the system to explore edges by real world locomotion of the user. It also shows that per-frame processing can augment perception more cost-effectively than conventional methods by creating tighter interaction loop. Also, this kind of real world interaction would be beneficial to hands-free implementations in the future.

Inaccuracy and slowness of line segment detection is found to be a limiting factor in the current implementation. This could be remedied by depth-guided segment detection or a fast GPU-accelerated implementation in conjunction with more sophisticated layout optimization.

**Acknowledgements.** This work was supported by JSPS KAKENHI Grant Number 23240016.

## References

1. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. *Multimedia Tools and Applications* 51, 341–377 (2011)
2. Aulinas, J., Petillot, Y.R., Salvi, J., Lladó, X.: The SLAM problem: a survey. In: *Catalonian Conference on AI*, pp. 363–371 (2008)
3. von Gioi, R.G., Jakubowicz, J., Morel, J.-M., Randall, G.: LSD: a Line Segment Detector. *Image Processing On Line* (2012)
4. Mistry, P., Maes, P.: Sixthsense: a wearable gestural interface. In: *ACM SIGGRAPH ASIA 2009 Sketches*. SIGGRAPH ASIA 2009, pp. 11:1–11:1. ACM, New York (2009)
5. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* 12(2), 1437–1454 (2012)
6. Adelson, E.H.: Lightness perception and lightness illusion (1999)

# Painting Alive: Handheld Augmented Reality System for Large Targets

Jae-In Hwang, Min-Hyuk Sung, Ig-Jae Kim, Sang Chul Ahn,  
Hyung-Gon Kim, and Heedong Ko

Imaging Media Research Center, Korea Institute of Science and Technology

**Abstract.** This paper presents a handheld augmented reality (AR) system and an authoring method which provides alive contents in large targets. In the general augmented reality tools, they are not designed for large targets but for only adequate size of target which fits in the screen. Therefore we designed and built a vision-based AR system and an authoring method that can handle much larger targets than the view frustum.

**Keywords:** augmented reality.

## 1 Introduction

Recently, handheld augmented reality (AR) technology is growing drastically with the advances of mobile devices such as smartphones and tablet-PCs. There are many diversities of application for handheld AR such as games, visual information providers, advertisements, and so on. One of the adequate applications is mobile tour guide in the museum or sightseeing places [1]. While working on our project named Mobile Augmented Reality Tour (MART) since 2009, we have found several interesting research issues about handheld AR. Because the goal of the project was providing augmented contents through mobile devices during tours in the museum or places, there were many targets that have various sizes and forms. In cases of small targets, we could apply existing augmented reality tracking algorithms or tools such as SIFT [2] or SURF [3]. But most of the tracking methods were designed for the screen-fit size object. Then what happens if we can see just ten percent of the object through camera in the mobile device? The system would have difficulties in recognizing and tracking the object. So the augmented contents would not appear or could be placed on the wrong position.

In the project MART, we had technical challenges to adding augmenting contents on the “Painting of Eastern Palace” which is 576 centimeters in width and 273 centimeters in height. The behavior of tourist using handheld AR tour guide is not predictable. They could focus on any certain part of the painting from various viewing positions and directions. So the technical challenge here was building handheld AR tracking system with the unpredicted view of the large target. In the remaining part of this paper, we will show details of the tracking system. Moreover, we also present about the authoring method for providing various multimedia contents for the objects.



**Fig. 1.** Example of large target for handheld AR, Painting of Eastern Palace (National Treasure of South Korea, located at the museum of Donga University, 576cm×273cm)

## 2 Vision-Based Tracking for Large Targets

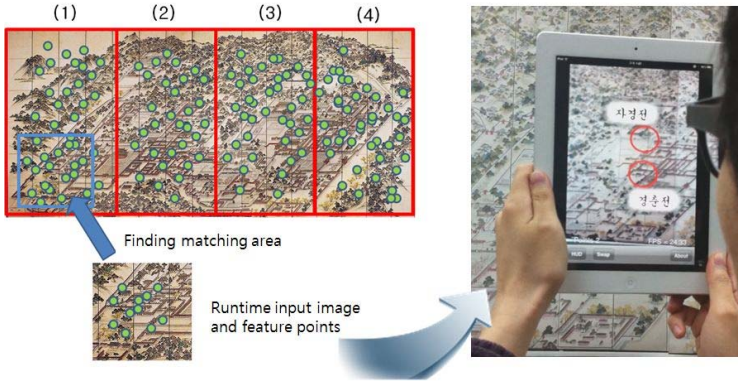
### 2.1 Divide and Conquer Method in Handheld AR

The basic idea of the large target tracking is simple. We divide the large target into multiple pieces and store feature points in the database. At the beginning of the execution of the AR program, we compare input features with features of database. We can find which part we are looking at by comparing feature points of each piece with input feature point set. This step is called target recognition. Once we find the piece what we looking at, we compute position and pose of the camera. The feature matching process is described in the Sec. 2.3. When we lost the target, we do the same procedure again. The time for finding target piece depends on the numbers and resolution of the pieces. In our case, we divided our target into four parts, and recognizing each piece takes less than 50 milliseconds. (If it takes more, the latency of the tracking would be noticeable.) So, it would take less than one second to recognize the current target and compute camera poses among 7 to 8 large targets.

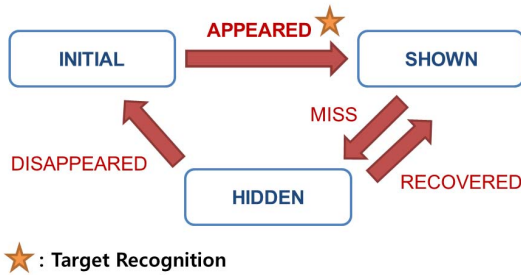
To avoid the failure of the matching caused by the change of user's viewpoint, we trained different scale of the target. This procedure produces more image pieces than just dividing. In our case, we added 4-5 more pieces in case when the user wants to see specific part of the painting. As the result, it took within one second to find the part when we lost them.

### 2.2 Tracking States

In cases of handheld AR, abrupt movement of handheld cameras becomes the main factor that makes the tracking procedure unstable. Particularly in our



**Fig. 2.** Feature matching method for large targets



**Fig. 3.** Tracking states and transitions between them

system, it can be triggered to find the new target piece even for one moment miss of the current target. The same situation of momentary target miss can also be occurred due to camera noises that are usually observed in low quality cameras used by mobile devices. In our system, we avoid this problem by defining three states in the tracking procedure as shown in Fig. 3. From the “Initial” state, we move to the state “Shown” when a target piece is found in the recognition step. When the found target disappear in the camera view, we do not directly move to the “Initial” state, but temporary move to a state “Hidden”. This “Hidden” state indicates that the current target momentary disappeared but will be shown soon. The target recognition step is executed only when we have moved back to the “Initial” state. This hidden-state system may delay the shift from one target piece to the other, but this artifact is merely noticeable if we define the lasting time in “Hidden” state as short. In our implementation, we set 5 frames for the “Hidden” state period.



Fig. 4. Example of 2D image strip for animation

	A	B	C	D	E	F	G	H	I	J	K
1	image	king	king	566	574	128	128				
2	image	red_men	redman_bow	196	788	150	150	10	0.02		
3	image	yellow_men	yellowman_bow	530	788	150	150	10	0.02		
4	image	spear_men_1	2spearman	838	470	186	186	3	0.1		
5	image	spear_men_2	3spearman	698	618	220	220	3	0.1		
6	image	spear_men_3	4spearman	236	546	254	254	3	0.1		
7	image	text_box_01	text_box	360	50	560	272				
8	text	scene_txt_01	scene_01	380	68	520	236	24	0.0005		
9	sound	scene_wav_01	scene_01								
10											
11											

Fig. 5. Spreadsheet style layout of contents

## 2.3 Matching and Tracking Method

We used modified version of Histogrammed Intensity Patch (HIP) to match and tracking in real time [4]. In the training step, we generated a set of training images by artificially warping a reference image with various scales, rotations, and other affine transformations. Local image patches are extracted from training images and grouped when they are obtained from close position with similar warping. In each group of local patches, we create a simple integrated patch which of each pixel is quantized into 5 levels using histograms. When, these quantized patches are also produced in runtime, it can be done much faster since we only handle one specific viewpoint. Moreover, the matching between patches can also be computed quickly using bit-wise operations.

## 3 Multimedia Contents Layout Authoring for Handheld AR

### 3.1 Multimedia Contents Types

There are many different types of digital contents which can be presented through handheld AR. In our system, we decided to show 2D animation which can be blended easily on the old painting. Also, we added audio/text narrations which can deliver historical stories.

2D animation can be made with series of images for each frame. As shown in Fig. 4, the images are stitched in one image. The animation is shown during run-time by putting each part of images onto the frame buffer.



Fig. 6. Result of the layout authoring

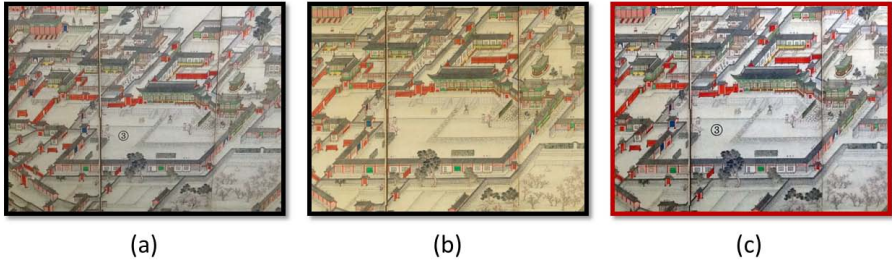
### 3.2 Layout Authoring

We use spreadsheet style layout for the contents authoring. Using the layout style, it is very easy and intuitive to locate and display various digital contents on the screen. In the Fig. 5, the first column represents the type of the contents. The second column contains object names and third column shows names of resource files. Also, the next 4 columns indicate the left upper position ( $x$ ,  $y$ ) and size ( $x$ ,  $y$ ), respectively. The last 2 columns are optional for animation; the number of strip frames and animation speed. In case of the text, we can animate text scrolling by putting animation speed at the column ‘T’.

Fig. 6 shows the result of the layout authoring. We placed scrolling textbox, animated characters and a narration sound. When the user looks the painting through the handheld camera, multimedia contents appear instantly. As described on the spreadsheet-style layout, king and other characters are located. The textbox shown in the figure is an animated textbox, therefore it scrolls itself as time goes on.

### 3.3 Illumination Adaptation

In some practical applications, we may experience a decline in the tracking performance due to illumination of real situations. For example of museums, the lighting is generally very dark for protecting displayed stuffs from being exposed to strong lights. It means that the camera captured image can look quite different with the trained image as both images cannot be recognized as the same one in the tracking procedure. A common solution for this case is to simply use the camera captured image from the training step. In this case, however, it is too difficult to place virtual objects in proper positions of the augmenting 3D space. As an example, it is practically impossible to capture the image from the exact front view. This indicates that the real image plane may be a bit tilted in the captured image as shown in Fig. 7 (a). Hence, we cannot put a virtual object to be exactly fit on the real image plane.



**Fig. 7.** Tone mapping. (a) A captured image, (b) The ground truth image, (c) The transformed image from (a) to (b)

To solve this issue, we focused on the trade-off between robustness and computation time of feature descriptors. Our modified HIP descriptor provides real-time speed, but cannot overcome the difference of illumination unless all diverse lighting conditions are considered in the training step. (Even if they were so, considering more conditions lead to slower speed in runtime.) On the other hand, some other descriptors such as SIFT that are too slow to be used in real-time applications may work robustly even for the difference of illumination. Therefore, as a pre-processing step, we transform the given captured image using SIFT [2] to be looked the same with the ground-truth image. Indeed, we perform this as a sort of tone mapping process. As shown in Fig. 7, the transformed image (c) not only has the exactly same arrangement of scenes with the ground truth (b), but also reflects the illumination of the real situation as (a).

## 4 Conclusion and Future Work

In this paper, we presented a handheld AR system which can show information for the very large target. We presented the method for the matching and tracking for large targets by divide and conquer. Also, we presented the spreadsheet style based layout authoring for AR contents. By building total procedure, we could show various augmented contents without much efforts. Developed system has been installed at the National Palace Museum of Korea. We have several future plans to improve current system. In the current system, we can detect a few but not many large targets such as more than hundred targets. We could do that by adding detection module which can detect numerous targets before tracking stage.

**Acknowledgement.** This research is supported by Ministry of culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA), under the Culture Technology (CT) Research & Development Program 2009.

## References

1. Vlahakis, V., Ioannidis, M., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., Gleue, T., Daehne, P., Almeida, L.: Archeoguide: an augmented reality guide for archaeological sites. *IEEE Computer Graphics and Applications* 22(5), 52–60 (2002)
2. Ke, Y., Sukthankar, R.: Pca-sift: a more distinctive representation for local image descriptors. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, June-2 July 2004*, vol. 2, pp. 506–513 (2004)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
4. Taylor, S., Rosten, E., Drummond, T.: Robust feature matching in 2.3  $\mu$ s. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 15–22 (June 2009)



# VWSocialLab: Prototype Virtual World (VW) Toolkit for Social and Behavioral Science Experimental Set-Up and Control

Lana Jaff<sup>1</sup>, Austen Hayes<sup>2</sup>, and Amy Ulinski Banic<sup>1</sup>

<sup>1</sup>University of Wyoming, Dept. 3315, 1000 E. University Ave, Laramie WY, 82071

<sup>2</sup>Clemson University, 100 McAdams Hall, Clemson, S.C. 29634

{ljaff, abanic}@uwyo.edu, ahayes@clemson.edu

**Abstract.** There are benefits for social and behavioral researchers to conduct studies in online virtual worlds. However, typically learning scripting takes additional time or money to hire a consultant. We propose a prototype Virtual World Toolkit for to help researchers design, set up, and run experiments in Virtual Worlds, with little coding or scripting experience needed. We explored three types of prototype designs, focused on a traditional interface with pilot results. We also present results of initial expert user study of our toolkit to determine the learnability, usability, and feasibility of our toolkit to conduct experiments. Results suggest that our toolkit requires little training and sufficient capabilities for a basic experiment. The toolkit received a great feedback from a number of expert users who thought that it is a promising first version that lays the foundation to more future improvements. This toolkit prototype contributes to enhancing researchers' capabilities in conducting social/behavioral studies in virtual worlds and hopefully will empower social and behavioral researchers by proving a toolkit prototype that requires less time, efforts and costs to setup stimulus responses types of human subject studies in virtual worlds.

**Keywords:** Virtual Humans, Online Virtual Worlds, Virtual Environments, Social Science, Psychology, Behavioral Science, Human Experiments, Toolkit, Evaluation, Prototype Experimental Testbed.

## 1 Introduction and Motivation

Online Virtual Worlds, such as second life and 3rd Rock Grid (3RG), have been widely used for educational and entertainment purposes [2,7,9,12]. Recently, these environments have also been sufficiently used as platforms to conduct studies and experiments in various fields [3,5,10]. Virtual Worlds have economic and political systems that provide interesting social dynamics that have been studied by researchers in social and behavioral sciences, yet not fully explored. Due to the computational nature of these virtual worlds, they lends themselves to be used as virtual laboratories for conducting human subjects experiments in the social and behavioral sciences, typically found in fields such as Sociology and Psychology, since many of the

environmental and social variables can be controlled, and automated data collection tools can be included. Over the years, social and behavioral researchers have shown great interest in using Virtual Worlds as platforms for conducting their studies [3,5,8,10]. Psychologists have expressed that virtual worlds increase participants' "engagement" and their reactions, therefore may increase the reliability and effects of the experiments on the participants [3]. Furthermore, virtual worlds may minimize the "lack of replication" and "sampling problems" found in traditional laboratories [2,7,10,13]. However, in previous studies conducted in Virtual Worlds, scientists had to either learn the scripting skills to set up an experiment or collaborate with other professionals, which may be time consuming and costly. The objective of this research is to design a method to empower researchers to be able to set up and conduct experiments in Virtual Worlds without needing to learn scripting and depend on other professionals. The goal of this research was to use virtual worlds as an alternative platform to design and conduct standard social and behavioral science experiments.

We designed and developed all the components needed to conduct an experiment in a Virtual World that involved a simple task, mixed design of between and within subject conditions, a controlled virtual character involved in the experiment, and multiple data types collected. Additionally we interviewed several social and behavioral science researchers to learn about their experimental goals and limitations. As a result, we propose a prototype Virtual World Toolkit for researchers to use to help design, set up, and run experiments in Virtual Worlds, with little coding or scripting experience needed. We explored three types of prototype designs: one visual with drag and drop features, one conversational where a virtual human discusses the experimental study and then implements it, and the third a more traditional button-like interface. Our hypothesis was that our toolkit is easy to use, has sufficient capabilities to set up and conduct a basic human subject experiment, and requires little training and less time, as opposed to learning scripting skills, to get started. We also hypothesized that our toolkit will receive positive attitude ratings similar or better than the other toolkits that the researchers currently use to conduct their studies.

## **2 Background and Related Work**

There have been a number of studies conducted in Second Life virtual world to evaluate the effects of the presence of virtual characters on task performance. In real world social and behavioral studies, there is a theory that refers to the effects of presence of others (real humans) during task performance called social facilitation/inhibition [4,8,14]. This theory states that the presence of others affects the performance of novel tasks more than the performance of learned tasks [4,8,14]. Participants perform better on learned tasks and worse on novel tasks. A number of studies were conducted in virtual environments (including online virtual worlds such as Second Life) and aimed to study the social facilitation/inhibition with the use of virtual humans as the audience while real participants perform different levels of tasks [4,8,13,14]. Hayes conducted a study to evaluate the social facilitation/inhibition effects among simple and complex tasks [8]. This study found that the social

facilitation/inhibition theory applies to virtual worlds where the results showed that the participants were affected by the presence of the virtual observers (male or female) during the performance of the complex tasks. Yee conducted another study to evaluate the effects of the avatar's gender, distance and eye-gaze on the social behaviors between avatars in virtual environments [11]. The results of this study showed that male avatars tend to keep a larger personal distance and less eye contact with other male avatars, while female avatars keep a smaller personal distance with other female avatars and more eye contact. Antonio investigated the connection between avatar's behaviors with one another in virtual worlds like eye contact, conversations and the application of "clustering techniques" [1]. The results were applied to real life relationships between teachers and students by better understanding the social behaviors of students in class. The data collected from the study showed that some students were paying attention, especially with eye contact, to lecture while others did not.

### **3 Preliminary Research: Exploring Components Needed to Conduct Social and Behavioral Studies in Virtual Worlds**

For our preliminary research, we determined the necessary components to conduct a simple stimulus-response human-subjects studies. We designed the toolkit to allow for flexible options for experiments to identify independent variables, such as the between and within subject conditions, and dependent variables, such as the automated data collected, trials, orders and animations, for automated virtual character control. The toolkit takes advantage of the scripting capabilities of the virtual world, yet abstracts that from the researchers by presenting buttons and menus to interact with rather than scripting the components. We developed the toolkit on a 3D Rock Grid leased island (3RG). Our prototype was based on a social-facilitation study with virtual human audience types. Researchers can use and interact with the toolkit using an avatar, or a computer generated character controlled by a human. The capabilities are:

- Number of conditions: Consists of an empty edit field where the number of conditions is entered using the keyboard.
- Manipulation of conditions: Select one of two options to manipulate the study conditions (within subjects, between subjects) by clicking on an option.
- Observer Avatar: Consist of three options as they relate to between subjects conditions. This data will be sent to the data collection note card for reference.
- Participant Avatar: Consist of two options: Male and Female. This data will be sent to the data collection note card for reference as they relate to the experimental participant that will take part in the study when researchers set up their experiments.
- Number of trials: Number of trials is entered using the keyboard.
- Order of trials: Select one of three options that represent the order in which the trials are presented in (randomize, Specific order, Balanced Latin squares) by clicking on an option using the left mouse button.

- Response time: Includes three options to select how to record the response time (per block, per trial and per condition) by clicking on an option.
- Completion time: Includes three options to select how to record the completion time (per block, per trial and per condition) by clicking on an option.
- Task accuracy: Defines how to measure accuracy of the tasks using the keyboard.
- Task errors: Defines what the errors of the tasks are using the keyboard.
- Input Method: What is used as an input method to respond to the tasks? Keyboard is the only input method provided for the toolkit.
- Task output: What is used as the output stimuli (to provide feedback for the user's responses)? Two options are provided: Textures and Play sound.
- Start: This is the last button to press in order to start the study after setting up the above components by clicking on this button using the left mouse button.
- Avatar Appearance: Provides four common and required appearances to conduct studies in virtual worlds: professional, casual, hot-trendy looks and bold (rock and roll) looks. For each of these appearances, we provided different body shapes (tall, short, skinny, muscular), skin colors (white,dark,tan,male base), eye colors (green,blue,brown,grey) and hair styles including bald.
- Avatar Gestures: Provides 10 common gestures for the observer avatar that observes the participant's avatar during task performance. These gestures include sad, angry, impatient, embarrassed, laugh, unhappy, wave, worried, cough and bored.
- A task display board: This board displays the tasks of the set up study for the participant after pressing the "Start" button by the researcher.



Fig. 1. Example of the prototype interface for parts of the toolkit in the Virtual World

## 4 Experimental Design

### 4.1 Physical and Virtual Environments

All participants completed experimental tasks and questionnaire in a physical testing room at L41 lab of the Engineering building at the University of Wyoming. We used an Intel Core 2, 2.66 GHz, 2 GHz RAM Dell PC with an ATI Radeon HD 160 graphics card attached to a 20 inch at screen monitor to display the virtual world. In addition, we used a different PC (Intel Core (TM) 2.80 GHz, 2 GHz RAM Dell PC with an NVIDIA Quadro FX 580 graphic card attached to a 20 inch at screen) for complet-

ing the questionnaires. Zen Viewer version 3.4.1 was used to view the 3RG virtual world. The participants were given a gender-appropriate avatar to control and interact with our toolkit. Our toolkit was created in a 3RG leased island. It consists of a collection of objects integrated together and scripted using Linden scripting language (LSL). A gender appropriate avatar (called the observer avatar) was given to the participant to use to setup studies using the toolkit. Another avatar was given to the participants (called the experimental avatar) and the gender was varied for this avatar. The experimental avatar was used to play the role of an experimental participant and perform that tasks of the studies set up in tasks A and C.

## 4.2 Experimental Tasks

Our experiment consisted of three tasks that were manipulated within subjects.

**Task A.** We asked the participants to set up the components of a specific stimulus-response type of study. We were assessing usability in this task. The participants were given specific math task components and were asked to set up the study using our toolkit without training learnability. For this task, we recorded the order in which the participants set up the study components as well as the number of presses used. After completing the tasks of this study, we showed the participant how to access and view the data collected from the study.

**Task B.** Participants created gestures and adjusted appearance for an avatar. The first level of task B included using our toolkit without training to modify the avatar's appearances according to specific criterion (color of skin, color of eyes, body type and outfit) that were varied for each participant. This task also included creating two different gestures for that avatar. For the second level of task B, we gave the participant 10 minutes and asked them to figure out how to change the appearance of the avatar into a specific appearance (specific color of skin, color of eyes, body type and outfit) without using the toolkit. Afterwards we gave the participant another 10 minutes to create two specific gestures for that avatar without using the toolkit as well.

**Task C.** We asked the participants to set up a study of their own using the toolkit after giving them a short training on how to upload their own tasks into the toolkit. We were assessing our toolkit's capabilities and usability to set up stimulus/responses types of studies in this task. We had asked each participant to prepare and bring at most five files (due to time constrains) of basic stimulus responses type of study that they have conducted or familiar with in JPEG format to complete this task. We gave the participants a short training session on how to upload these tasks into the toolkit then asked them to set up the components of that study.

## 4.3 Experimental Measures and Procedure

Pre-experimental questionnaires collected demographic data (gender, age, ethnicity, computer and virtual worlds use level) and experimental background data. We asked participants to rate the usability and user experience criterion of the toolkits that they

previously or currently use to conduct human subject studies and were assessed on a 7-point numerical scale (1=Not at all to 7= a great deal). The learnability questionnaire was given to the participants after completing task A. We also asked participants to complete a questionnaire asking them questions about their attitude and opinions towards other toolkits (after task B, part 1) and towards our toolkit (after task B, part 2). We assessed the attitude responses on a 5-point Likert scale (1=strongly agree, 5=strongly disagree). Item responses on a final questionnaire (after task C) were used to determine whether our toolkit has sufficient capabilities to perform appearance modification and create gestures for the avatar in an easier, less time consuming method as appose to the traditional methods in these environments. This questionnaire collected responses on a 7-point numerical scale (1=Not at all to 7= a great deal). We also asked open-ended questions about the toolkit and recommendations. Participants also completed a post-experimental co-presence questionnaire, which refers to the extent the participants felt they were inside the virtual world and interacting with the avatars, on a 7 point numerical scale.

Prior to the experiment, we gained consent from each participant and asked them to complete the pre-experimental questionnaire. Participants were given brief training on how to use the avatar to move and interact with objects. After, we provided the list of components of the sample experiment and asked the participant to set it up using the toolkit. After completing task A, the participants completed the learnability questionnaire then proceeded to task B. The participants were asked to perform specific modifications on the observer avatar's appearance as well as creating gestures for it, then completed questions about the toolkit. For the second level of task B, the same observer avatar is used by the participants to modify its appearance, create a gesture for it without using the toolkit, and answer questions about the toolkit. The order of the task B parts was balanced. After completing all three tasks, the participants were instructed to move to complete the co-presence, usability and user experience questionnaires. Finally, the participants were debriefed and thanked for participation.

## 5 Evaluation Results

Mean (M) and standard deviation (SD) of learnability, usability and user experience percentages were computed by averaging across grouped questions for each participant in the pre and post experimental questionnaires. Order and number of presses of each button were computed by summing each item in the observer check sheet which is a list of observations while participation sheet that was used to record the quantitative data represented in the order and times each button in the toolkit was pressed. A paired samples (t-test) was conducted to test difference of means in comparing our toolkit against other applications used for conducting human subject studies and usability for virtual character control, where  $p = 0.05$  was used to indicate significance. There were 6 expert participants, Faculty, PhD. and master students who conduct subject studies in the Psychology and computer science departments (however, not familiar with virtual worlds), males and females, and from the University of Wyoming. The mean age for the participants was 29.5,  $SD = 4.1$  and they were randomly assigned a gender appropriate avatar.

## 5.1 Sufficient Capabilities

Our participants found that the toolkit meets the needs to set up standard stimulus responses types of studies, where ( $M = 5.3$ ,  $SD = 1.3$ ) on a scale of 1-7. Results showed that the toolkit was rated as sufficient and high to set up stimulus responses types of studies, where ( $M = 6.5$ ,  $SD = 0.54$ ). We asked the participants to write down the capabilities that our toolkit provides to set up and conduct stimulus response type of human subject studies in the post experimental questionnaire. These capabilities include in allowing as many conditions as necessary, allowing the setup of the study's tasks, trials and orders, as well as allowing multiple data collection tools and output stimuli. The results showed that the toolkit provides enough capabilities to setup stimulus response types of studies in virtual worlds where ( $M = 5.3$ ,  $SD = 1.10$ ). After calculating the mean and standard deviation for future usage and recommendation, the results showed that the participants are more likely to use the toolkit to set and conduct their future studies if they were to conduct studies in virtual environments and that they will recommend it to their peers. Where ( $M = 6$ ,  $SD = 0.89$ ) for likelihood of using the toolkit in the future studies in virtual environments and ( $M = 5.6$ ,  $SD = 1.21$ ) for recommending the toolkit to others.

## 5.2 Learnability, Usability and User Experience

The toolkit was rated positively by participants in regards to learnability, where  $M = 6.21$  and  $SD = 0.49$ . Many users reported that they did not need programming/scripting experience to use the toolkit was calculated across participants, where  $M = 7$ ,  $SD = 0$ . The results show that the participants did not need a lot of support to set up studies using the toolkit, where ( $M = 5.33$ ,  $SD = 0.81$ ). The results of a paired samples t-test to determine change in attitude from level 1 to level 2 of task B, showed that it is significantly easier to learn how to modify the avatar and create gestures using the toolkit without training, where  $t(5) = 8.216$ ,  $p < 0.001$ ,  $M = 4.60$ , and  $SD = 0.51$  and  $t(5) = 23$ ,  $p < 0.001$ ,  $M = 4.80$  and  $SD = 0.40$  respectively, compared to learning how to modify the avatar and creating gestures using the traditional methods, without training, where ( $M = 1.6$ ,  $SD = 0.81$ ) and ( $M = 1$ ,  $SD = 0.00$ ) respectively.

As expected, the results show that setting up studies using the toolkit saves a significant amount of time, where ( $M = 6$ ,  $SD = 0.89$ ). The results show that the participants rated the toolkit as highly intuitive to set up stimulus responses types of studies where ( $M = 6.16$ ,  $SD = 0.40$ ). The participants rated the toolkit high in terms of consistency between other applications where ( $M = 6$ ,  $SD = 0.63$ ). The results of a paired samples t-test to determine change in attitude from level 1 to level 2 of tasks B, revealed that it is significantly easier to modify the avatar using the toolkit  $t(5) = 19$ ,  $p < 0:001$ , where ( $M = 5.0$ ,  $SD = 0.0$ ) compared to using the traditional virtual environments methods where ( $M = 1.8$ ,  $SD = 0.16$ ). It saves a significant amount of time to modify the avatar using the toolkit as expected, where  $t(5) = 6.32$ ,  $p = 0.001$ ,  $M = 4.50$  and  $SD = 0.54$  as opposed to using the traditional methods ( $M = 1.8$ ,  $SD = 0.75$ ). It is significantly easier to create gestures using the toolkit where  $t(5) = 10$ ,  $p < 0.001$ ,  $M = 4.60$  and  $SD = 0.81$  compared to the traditional methods, where ( $M = 1.3$ ,  $SD = 0.51$ ). It was also found that creating gestures for the avatar using the toolkit saves a

significant amount of time  $t(5) = 5.94$ ,  $p = 0.002$ ,  $M = 4.6$  and  $SD = 0.51$  if compared with the traditional methods of creating gestures in these environments, where ( $M = 1.8$ ,  $SD = 0.75$ ). The results showed that the participants rated the toolkit highly manageable to set up stimulus responses types of studies where ( $M = 6$ ,  $SD = 0.63$ ). The participants were satisfied with the toolkit, where  $M = 5.8$  and  $SD = 0.58$ . The results showed that that participants were significantly satisfied with the toolkit, where  $t(5) = 8$ ,  $p < 0.001$ ,  $M = 4.50$  and  $SD = 0.83$  compared to their satisfaction ratings on the traditional methods in virtual worlds, where  $M = 1.80$  and  $SD = 0.40$ .

### 5.3 Researchers Opinions: How Our Toolkit Compared to Other Similar Applications

A paired samples t-test results show no significant difference in comparing the toolkit's simplicity and ease of use to set up studies compared to other applications:  $t(5) = -2.291$ ,  $p = 0.071$ , where ( $M = 5.5$ ,  $SD = 0.54$ ) for the toolkit and ( $M = 3.3$ ,  $SD = 2.16$ ) for other applications. After comparing the learnability of the toolkit compared to other applications, it was found that the toolkit is significantly easier to learn without training  $t(5) = -2.557$ ,  $p = 0.051$ , where ( $M = 6.3$ ,  $SD = 0.51$ ) for the toolkit while ( $M = 3.5$ ,  $SD = 2.34$ ) for other applications. The results show that setting up studies using the toolkit saves a significant amount of time,  $t(5) = -4.503$ ,  $p = 0.006$ , where ( $M = 6$ ,  $SD = 0.89$ ), compared to other applications ( $M = 2.83$ ,  $SD = 1.83$ ), and significantly more sufficient to set up studies,  $t(5) = -4.108$ ,  $p = 0.009$ , ( $M = 6.5$ ,  $SD = 0.54$ ) as opposed to other applications ( $M = 3.5$ ,  $SD = 1.87$ ). The results show a significant difference in comparing the intuitiveness and manageability of the toolkit to set up studies compared to other applications where ( $M = 6.16$ ,  $SD = 0.40$ ) (for the toolkit),  $t(5) = -3.955$ ,  $p = 0.011$  and (for other applications)  $M = 2.66$ ,  $SD = 2.16$ ) and ( $M = 6$ ,  $SD = 0.63$  (for the toolkit),  $t(5) = -3.162$ ,  $p = 0.025$  and (for other applications)  $M = 3.33$ ,  $SD = 1.96$ ) respectively. Most of them made suggestions to enhance the toolkit and make it more suitable for their individual studies.

- "Brilliant idea for scientists in our field. I appreciate that I do not need to code".
- "The design is similar to software I usually use and I don't think I need a manual".
- "Excellent idea which has the potential to make research easier in our field. I mostly liked the design which was pretty obvious and easy to use".

## 6 Discussion

The results show that our participants rated the toolkit to be practical to use. The results show that the toolkit provides enough capabilities to set up stimulus responses types of studies. Not all participants agreed to use the toolkit for all their current or future studies in general. However, all participants concurred to use the toolkit for current and future studies conducted in a virtual world, and that they would definitely recommend the toolkit to their peers. This fulfills the objective of our toolkit to be used as an alternative platform to setup experiments in virtual worlds. The participants were able to take advantage



of the familiarity of the design with similar features to other applications that they usually use. The simple design of the toolkit has led the participants to complete the tasks for the first time without training. Many participants believed that they did not need scripting background or skills to use this toolkit to set up and conduct studies with accomplishes our objectives in creating a toolkit that is easy to learn and use with no or little scripting knowledge in the virtual world. The toolkit rated as significantly easier to use for avatars' appearance modification and gesture creation, than traditional methods. Our toolkit is limited to setting up stimulus responses types of tasks. Participants would like to add more capabilities, such as including adaptive types of task and response scales. In general, the toolkit meets the learnability, user experience and usability criterion. A number of expert users who thought that it is a promising first version that lays the foundation to more future improvements in order to make it more appropriate for setting up individual and more complex studies in virtual environments. The results show that the participants are more satisfied with our toolkit than others, though the researchers' opinions may be influenced by other factors that do not exist in our toolkit.

## 7 Conclusions, Contributions and Future Work

The results of this research have shown that the toolkit is easy to learn and use. It provides sufficient capabilities to setup stimulus response types of studies. The toolkit requires minimum training and coding skills and does not take too long to setup studies as opposed to learning the scripting skills. The results also showed that the toolkit provides alternative avatar control methods. The toolkit meets the learnability, user experience and usability criterion. The toolkit received a great feedback from a number of expert users who thought that it is a promising first version that lays the foundation to more future improvements in order to make it more appropriate for setting up individual and more complex studies in virtual environments. This toolkit prototype contributes to enhancing researchers capabilities in conducting social/behavioral studies in virtual worlds. The toolkit hopefully will empower social and behavioral researchers by proving a toolkit prototype that requires less time, efforts and costs to setup stimulus responses types of human subject studies in virtual worlds.

In the future, we plan to perform more extensive development of toolkit and conduct a more extensive study with more participants after more capabilities are developed. More future implementations would consist of including more advanced features for the toolkit such as including adaptive types of task (the answer to the previous question effects the next question), response scales and the ability to respond to the tasks with more than just yes and no. We will conduct a direct comparison between this toolkit and other similar applications. These features will provide more capabilities for the toolkit to setup and conduct more complex studies in virtual environments rather than limited to stimulus responses types of studies.

**Acknowledgements.** Thank you to participants of this research and who provided valuable feedback.

## References

1. Friedman, D., Steed, A., Slater, M.: Spatial social behavior in second life. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 252–263. Springer, Heidelberg (2007)
2. Gregory, B.E.A.: How are Australian higher education institutions contributing to change through innovative teaching and learning in virtual worlds? in changing demands, changing directions. In: Ascilite Hobart, pp. 475–490 (2011)
3. B. J.-B. A. C. Loomis, J. M.: Virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, and Computers* 31, 557–564 (1999)
4. B. R.-M. D. Sanders, G.S.: Distraction and social comparison as mediators of social facilitation effects. *Experimental Social Psych.* 14, 291–303 (1978)
5. Moschini, E.: The second life researcher toolkit: an exploration of in world tools, methods and approaches for researching educational projects in second life. Springer (2010)
6. Antonio Gonzalez-Pardo, E. P., de Borja Rodriguez Ortiz, F., Fernandez, D. C.: Using virtual worlds for behaviour clustering-based analysis. In 2010 ACM Workshop on Surreal Media and Virtual Cloning, pp. 9–14 (2010)
7. Lim, J.K.S., Edirisinghe, E.M.: Teaching computer science using second life as a learning environment. In: ICT: Providing Choices for Learners and Learning, p. 978 (2007)
8. Hayes, A.L., Ulinski, A.C., Hodges, L.F.: That Avatar is looking at Me! Social In-hibition in Virtual Worlds. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 454–467. Springer, Heidelberg (2010)
9. M. N. H. S. Zhang, Q.: A case study of communication and social interactions in learning in second life. In: 43rd Hawaii International Conference on System Sciences, p. 19 (2010)
10. Shailey Minocha, M.T., Reeves, A.J.: Conducting empirical research in virtual worlds: Experiences from two projects in second life. *Virtual Worlds Research*, vol. 3
11. Nick Yee, M.U.F.C., Bailenson, J.N., Merget, D.: The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology and Behavior* 18, 115–121 (2007)
12. S. E., Wiecha J., Heyden R, M. M.: Learning in a virtual world: experience with using second life for medical education. *J. Med. Internet Res.* 12 (2010)
13. Zanbaka, C., U. A.,G. P., L. F. H.: Effects of virtual human presence on task performance. In: International Conference on Artificial Reality and Telexistence (ICAT) (2004)
14. Zanbaka, C., Ulinski, A., Goolkasian, P., Hodges, L.F.: Social responses to virtual humans: Implications for future interface design, pp. 1561–1570. ACM Press (2007)

# Controlling and Filtering Information Density with Spatial Interaction Techniques via Handheld Augmented Reality

Jens Keil<sup>1</sup>, Michael Zoellner<sup>2</sup>, Timo Engelke<sup>1</sup>,  
Folker Wientapper<sup>1</sup>, and Michael Schmitt<sup>1</sup>

<sup>1</sup> Fraunhofer IGD, Darmstadt, Germany

<sup>2</sup> Hof University, Germany

**Abstract.** In our paper we are proposing a method for contextual information filtering based on the user's movement and location in order to enable the intuitive usage of an "internet of things" via augmented reality (AR) without information overload. Similar to Ray & Charles Eames' "Power of Ten" and Jef Raskin's "Zooming Interface" we are displaying seamless information layers by simply moving around a Greek statue or a miniature model of an Ariane-5 space rocket. Therefore we are employing concepts of camera- and motion-based interaction techniques and use the metaphors of "investigation" and "exploration" to control the way augmented and visually superimposed elements are presented in order to mediate information in an enhanced and engaging manner with aspects of digital storytelling techniques.

**Keywords:** Adaptive and personalized interfaces, Human Centered Design, Information visualization, Interaction design, New Technology and its Usefulness.

## 1 Introduction

The idea of retrieving and viewing geo-located information with augmented reality (AR) additionally or compared to traditional maps has been eased with the advent of handheld augmented reality on recent smartphones. AR browsers like Junaio, Layar, Nokia World Lense and other apps superimpose geo-referenced information on top of the camera's video stream like annotations, which relate and stick to a point/target-of-interest in the environment. Usually, these annotations are grouped in special "channels" and are visually depicted by icons or textual labels and are in reference to a specific place or position. Although oftenly used, they aren't of course technically limited to geo-referenced information.

With this video-see-through effect on smartphones, it has more generally become a commodity to superimpose annotations in AR views: technically speaking, in contrast to creating rich 3D models and assets, annotations are quite informative, they do not need intense processing power of GPUs, and their creation is relatively simple and of low costs. However, adding such descriptive information obviously come with some visual drawbacks: those annotations and labels

may occlude real object or more seriously a point-of-interest in the video or, if not aligned accordingly, they may tend to create ambiguity, since it might not be clear, to which target the information belongs. Moreover, if not filtered, the amount of too many augmented information turns the idea of data/information exposition and clarification into a visual overload of the scene.

Inspired by these issues and with the question of how to present and ease data access in handheld augmented reality, the methods we are going to present in this paper take on the concepts of motion controlled interaction in AR, and of zooming interfaces, as e.g. proposed by Raskin [6], where users control the density of information presented by scaling the viewed area in order to see more or less details while browsing through data. With our approximation technique we transfer this concept from originally virtual 2D spaces into application in real 3D environments, where the density and shape of augmented and overlaid information changes according to the distance from device to tracking target. We believe that this approach will not only make the experience of contextual information more engaging and intuitive but also more lasting, because users may focus on perceiving relevant information instead of coping with the user interface (UI) structure of the application that presents it.

## 2 Related Work

Almost all handheld AR apps nowadays work like magic lenses. The term was introduced at Siggraph 1993 by [2]. They proposed a new see-through interface to "reveal hidden information, to enhance data of interest, or to suppress distracting information". Originally designed for display interfaces this concept is to be found in most augmented reality applications today. Up until now, there exist several methods to dynamically adapt the content richness of the augmented data, especially, when dealing with labels and annotations. In handheld AR apps representations appear usually rather static: targets such as magazine pictures or posters are extended with text, videos or simple 3D assets (cf. IKEAs catalogue app, [10]). Interaction, if any, are click targets that link to external content. In case of AR-browsers, which are mainly used while being on the walk, only the user's movement, i.e. position, device heading, and orientation influences changes of the visual representation of data.

The work of [8] presents concepts for automatised positioning of superimposed annotations to avoid cluttering and occlusion in an adaptive manner. In terms of information flow, however, the presented methods are quite global and aren't controlled by the user itself, though. Although they tend to structure augmented elements, they do not introduce an informed order or sequence of data elements.

With adaptive visual aids [9] presents a technique, where the guidance level of superimposed annotations, i.e. the strength of information presented, changes dynamically. This mainly results in different visualisations, starting from unobtrusive visual indicators up to complex animations. The presented methods are explicitly tailored to support maintenance tasks. Users may not influence directly the depth-level, or eventually need to click through a menu beforehand,

which would interrupt and interfere with the video-see-through experience. The technique also doesn't account the absolute or relative position from user and device to the target-of-interest.

The idea of mapping gestures and well-known interaction tasks (such as selection, scrolling, navigation and object manipulation) to motion has been closely discussed in literature ([12] - [15]). The work presented in [5] and [4] uses camera and motion-based interaction in a way to control application and content paradigms that account camera-acquired targets in reality and device motion to turn the handheld device into a versatile interface. However, the view of interaction for both is technology-driven: [5] illustrates scenarios where users may "select" functions depending on their position to the tracking target by moving the device from far-to-near (fly-through) in order to e.g. de-/re-activate clipping masks that let users look inside an superimposed 3D model. Although their concept works with the camera/device position, the concept mainly maps rather arbitrary functions to movements in order to trigger application-modes, which might possibly work better with traditional menus. [4] proposed a conceptual framework of movement patterns, which are connected to a printed marker-map that allows to retrieve rotation or distance from device to target with computer vision. In order to use them accordingly, users need training in order to learn the basic gesture/movement-set and moreover the more complex ones.

### 3 User-Centric Spatial and Motion Interactions in AR

With computer vision based 2D and 3D target-tracking as a core enabling technology, the concepts that are going to be presented in this paper underlie paradigms and interaction-methods that are user-centric, and thereby in contrast to the typical understanding of gestures and patterns in computer science.

Instead of having pattern-sets, our idea is to link interaction to the superimposed visuals in a way that feels natural. We are using movements and mental models that are familiar to the user and thus almost instinctively applied. By doing so, we connect superimposed annotations and other content to motion based interaction and to camera-acquired targets, where the paradigms do not necessarily need to be trained beforehand. This connection should be perceived as joyful, easy-to-understand and passively, since it influences, filters and re-arranges superimposed material, but doesn't act as a change between system modes.

Being also based on the core paradigm of handheld augmented reality, where devices behave like a magic lens [2] (deployed through the video-see-through effect) our "gestures" make use of the natural movements that users already do, when working with handled augmented reality. Thereby the gestures we present work like AR-gazing, which is pointing extendedly at the same area, as well as observing, which correlates with walking around a target-of-interest, and inspecting and magnifying, which uses the distance/proximity to a target to influence the level-of-depth of information presented.

### 3.1 Conceptual Outline of Interaction Principles

While working with AR on mobiles, we have explored that users intuitively tend to align a focus or point of interest in the middle/center of the screen. Hence, as a core metaphor, we employ this as a gazing technique for usual selection tasks, where an extended pointing over time acts as a trigger in order to select virtual superimposed elements or ones from reality. We can think of two ways to implement this: while in the first one the trigger is represented visually by a small crosshair and thereby close to the representation of a mouse pointer (cf. fig.1), the second one lacks such an obvious representation and is displayed much more passively. Nevertheless, both are solely controlled by device movement.

In an extended and more elaborated way, the system intentionally adapts content information density (cf. [1]) of visual presentations to the user's location and distance to the object of interest: the greater the distance to the object the lesser the information to be displayed, based on a priority sequence. When moving closer to the object new information entities are appearing while other eventually irrelevant ones will disappear. This concept works similar to Eames' 1968 documentary short film "Powers of Ten" [3] where magnitudes are illustrated by a flight from a picnic on earth out to the outer edges of the universe and back into a proton of a carbon atom in a blood cell with the speed of ten times magnification every ten seconds. It also continues Jef Raskin's idea of zooming interfaces [6] by replacing the mouse with a smartphone held by a moving user and adding the dimension of viewing angles.



**Fig. 1.** AR-Gazing: an embodied pointer acts as a visual proxy, where in reality selected components are virtually highlighted



**Fig. 2.** Image sequence showing enhanced interaction coupled to approximation, where information density increases or decreases accordingly

Additionally, a second parameter for locative information adaption is the user's angle to the object: certain digital information is connected to a point on or surround the three dimensional object, and only visible from the right angle; one, where it makes sense to these information chunks to be consumable, while all other information is faded out.

### 3.2 Adaptivity

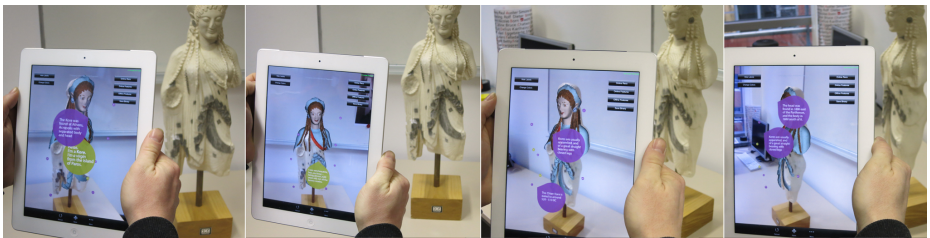
Information presentation and filtering is not considered as a binary process. A parametric model of thresholds allows to define hotspot-zones for each superimposed item. It reacts to the user's and the device's angle and distance alike and controls, whether the items appear or vanish from the viewpoint.

The result is a highly dynamic information space that encourages the user to explore it. Especially in museums and education, exploration is an important didactic principle, where mobile devices and AR can have strong impact on museum visitors: [16] could show that AR interfaces were perceived being intuitive with strong impact on user's memory and learning curve. Via AR, participants recalled information afterwards more reliable and seamlessly shifted their focus from digital to physical information and vice versa without any major problems.

## 4 Paradigms in Practice

We tested and implemented the paradigms in context of several prototype AR apps on iPads. We used our own instantAR framework [11], which runs on an iPad 4. The system tracks 3-dimensional tracking targets at roundabout 20 to 30 Hz, which makes the experience fast and stable enough in order to work with intensive device and user movement. Tracking is done with computer vision using an advanced version of [17] which is based on KLT and SLAM.

We applied our methods in several scenarios. At first, we experimented with pointing on a pump machine. Here the user could point and focus on a mechanical component, that was then visually highlighted (cf. figure 1). Once selected, additional information popped up on the screen. We implemented this scenario with and without a visual proxy.



**Fig. 3.** Image sequence showing locative information adaption depending to the user's angle to the tracking target

Within the second case we used the approximation principle in order to observe a physical miniature model of an Ariane-5 space rocket in an exhibition at Cit de l'Espace in Toulouse. While the user holds an iPad at a fair distance, the first level of superimposed overlays presents a lineup of different types of former or foreign rockets in order compare them in size and shape. The closer the user gets, the more detailed the augmentation becomes, where on a second level labels appear and inform about the structure, and enable to even look virtually into the interior on a third level; literally turning the mobile into an interactive x-ray device.

In a third case, we use device orientation and motion to control the amount of annotated information superimposed on exhibits at a time: digital audio and visual elements, which are spread around a real physical object, are shown or hidden depending on the angle of user/device to the real object in focus (cf. figure 3). The technique turns AR's "what's around me" into "whats in front of me", where the device is not only a virtual lens but also a pointer: just like gazing, a user may center or focus in order to select digital content on through devices camera. Therefore we used statues from the archaic gallery of the Acropolis Museum. Besides influencing the displayed elements, the fashion of soundness and speech changes accordingly: while the user is literally face-to-face to the statue, information is told in first-person, whereas being aside shifts narration to third-person.

Technically, each tracking target has a predefined front or ground axis. The distance of each superimposed item is calculated relatively to the core target. In relation to it, we calculate an entrance region for all items, which controls their disappearance and reappearance once the user approximates. By using these transitional zones, we are able to fade in and out each label depending on how close it is to the user's current view. It is important to mention, that each virtual item is view aligned and always fronting towards the user's view, which is done for readability reasons. With these zones we are not only able to control the general fading while the user walks around the target, we also may observe and control the relative distance from user to virtual items alike.

In order to be visually convincing without cluttering the scene we took care that the render process is able to deal with occlusion (meaning that the real physical object masks superimposed items) and that it handles back-face culling, even when an object is rendered transparently. By doing so, superimposed objects seamlessly fit into the video image and appear in the right depth order (and are this way perceived as being in front or behind the physical object). We thereby may also highlight regions of the real three-dimensional objects.

## 5 Findings and Informal Evaluation

In order to get an initial user feedback and to test the proclaimed paradigms, we conducted an informal user testing. A group of 15 people, all familiar with handheld augmented reality applications and computer vision tracking, have been asked to use the system with no more information than the introduction of



each tracking target. It was of interest to us to see, how each participant would start to use the system, how they would work with the paradigms, and if they would start understanding the principles just by using them.

To start, the biggest finding was, that all participants in general liked the idea of working with motion controlled interaction and that they understood the principles without major issues. In fact, the majority almost expected a movement-coupled interaction, which seems to be a key appeal of augmented reality, since it is the motion that reveals that virtual annotations stick to reality.

The gazing as well as the approximation principle were generally inconspicuous. Although we wanted to work without the visual pointer at first, it turned out that people found it rather confusing without that kind of feedback, which frankly worked as an indicator. Without it, people didn't understand exactly, why or how the system reacted and why it started to highlight mechanical components and giving extra annotations. However, with the proxy, some people were then trying to touch and move it with a finger, too.

The angle-dependent technique turned out not being as "intuitive" as expected: almost all participants tried to tab/click presented elements in order to "activate" content. Even more, people seemed inclined to be lazy in movement: 10 of the 15 were moving the device but did neither reposition themselves much nor really walk around the statue. Instead, at first they pointed (gazed) the iPad towards the augmented annotations, and eventually tried to tab them. Only, when explained, they started to move around and experiment with the technique. This seems, as if people are eventually not familiar enough or simply not used to the technique.

Finally, the approximation controlled depth-level filtering wasn't as "unexpected": people more or less instantly started to come closer or go away, but again weren't much moving around sideways. This might be coupled also to the overall situation and the way objects are exposed: the space should invite people and encourage to move around an object, too. This could also imply further design considerations, e.g. to make this interaction more recognisable through the UI in general. We also tend to recommend to design such interactions redundantly and combine gazing and tabbing additionally to angle-dependant control or use visual hints and animations that tease this kind of principles a bit.

## 6 Conclusion and Future Work

In this paper we presented a method for contextual information filtering based on the user's movement and location for augmented reality on mobile devices. Based on the magic lens paradigm of handheld AR, we introduced gestures that employ users natural movements, such as gazing (pointing extendedly at the same area), observing (walking around a target-of-interest) and inspecting and magnifying (distance/proximity to a target) to influence the level-of-depth and overall appearance of contextual and spatial presented information. The results of the informal evaluation were surprising and promising. We think there is still potential for a deeper quantitative and qualitative evaluation combined with

usability testings. Especially when including users without AR knowledge which takes large parts of museum's target audience and mass users alike. These tests should be carried out together with museum partners and their expertise.

**Acknowledgements.** This work is part of the EU-funded project CHESSE. The project aims to investigate personalised digital storytelling aspects in heritage applications that are tailored to interests of visitors by a user-centered and personalised design in order to enrich a visit of museums or science centers. Being close to rich narrations like presented in [7], the project explores not only location- and object-centered story telling techniques but uses also mixed and augmented reality.

## References

1. Tufte, E.: Beautiful Evidence. Graphics Press, Cheshire (2006)
2. Bier, E., Stone, M., Pier, K., Buxton, W., Derose, T.: Toolglass and magiclenses: The see-through interface. In: Proc. Siggraph 1993, Computer Graphics Annual Conference Series 1993, pp. 73–80 (1993)
3. Eames, C., Eames, R.: Powers of Ten (1968), <http://www.powersof10.com/film>
4. Rohs, M., Zweifel, P.: A conceptual framework for camera phone-based interaction techniques. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 171–189. Springer, Heidelberg (2005)
5. Harviainen, T., Korkalo, O., Woodward, C.: Camera-based interactions for augmented reality. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology 2009 (ACE 2009) (2009)
6. Raskin, J.: The humane interface: new directions for designing interactive systems. ACM Press/ Addison-Wesley Publishing Co., New York (2000)
7. Wither, J., Allen, R., Samanta, V., Hemanus, J., Tsai, Y., Azuma, R.: The Westwood Experience: Connecting Story to Locations via Mixed Reality. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (2010)
8. Grasset, R., Langlotz, T., Kalkofen, D., Tatzgern, M., Schmalstieg, D.: Image-driven View Management for Augmented Reality Browsers. In: Proceedings of 11th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2012 (2012)
9. Webel, S.: Multimodal Training of Maintenance and Assembly Skills Based on Augmented Reality
10. Inter IKEA Systems B.B.: IKEA Katalog App (2012 /2013), <https://itunes.apple.com/de/app/ikea-katalog/id386592716>
11. Engelke, T., Becker, M., Wuest, H., Keil, J., Kuijper, A.: MobileAR Browser A generic architecture for rapid AR-multi-level development. Journal Expert Systems with Applications 40 (2013)
12. Rekimoto, J.: Tilting operations for smal screen interfaces. In: Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (1994)
13. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C., Want, R.: Squeeze me, hold me, tilt me! An Exploration of Manipulative User Interfaces In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1998)

14. Hinckley, K., Pierce, J., Sinclair, M., Horwitz, E.: Sensing Techniques for Mobile Interaction. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (2000)
15. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M.: The Video Mouse: A camera-based multi-degree-of-freedom Input Device. In: Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (1999)
16. Damala, A., Cuband, P., Bationo, A., Houlier, P., Marchal, I.: Bridging the Gap between the Digital and the Physical: Design and Evaluation of a Mobile Augmented Reality Guide for the Museum Visit. In: Proceedings of DIMEA The International Conference on Digital Media in Entertainment and Arts (2008)
17. Wientapper, F., Wuest, H., Kuijper, A.: Reconstruction and Accurate Alignment of Feature Maps for Augmented Reality. In: IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (2011)

# Development of Multiview Image Generation Simulator for Depth Map Quantization

Minyoung Kim<sup>1</sup>, Ki-Young Seo<sup>2</sup>, Seokhwan Kim<sup>3</sup>,  
Kyoung Shin Park<sup>4</sup>, and Yongjoo Cho<sup>5</sup>

<sup>1</sup>Department of Computer Science, Sangmyung University, Korea

<sup>2</sup>Department of Computer Science, Dankook University, Korea

<sup>3</sup>Department of Computer Science, University of Tsukuba, Korea

<sup>4</sup>Department of Multimedia Engineering, Dankook University, Korea

<sup>5</sup>Division of Digital Media, Sangmyung University, Korea

pupleshine@gmail.com, windzard@empal.com,  
seokhwan@live.com, kpark@dankook.ac.kr, ycho@smu.ac.kr

**Abstract.** This study presents the novel multiview image generation simulator system based on the Depth Image-Based Rendering (DIBR) technique. This system supports both actual photographs and computer graphics scenes. It also provides the simple plug-in for pre-processing of depth map or post-processing of hole-filling algorithm. We intended to make this system as a platform to conduct various experiments such as the number of cameras, a depth map precision, etc. In this paper, we explain the design and the development of this simulator and give a brief comparative evaluation on linear and non-linear depth quantization method for computer graphics 3D scenes. The results showed that non-linear depth quantization method produced better performance on 7- to 3-bit depth levels.

**Keywords:** Depth Image Based Rendering, Multiview System, Depth Map Quantization, Hole-Filling.

## 1 Introduction

Recently three dimensional contents, displays and systems became more common and popular after the success of the three dimensional movie “Avatar”. After that, several major display manufacturers released commercial stereoscopic 3D TV or projection display products to the mass market. A lot of 3D movies have also created or reconstructed from 2D movies. These emerging 3D contents and display technologies attract public attention due to high fidelity realism and immersion. However, stereoscopic displays require users to wear stereoscopic glasses, which is inconvenient and cumbersome for the long-term uses. On the other hands, autostereoscopic display technologies, such as holography, volumetric display, integral imaging and multiview systems, are more comfortable as they do not require the use of special glasses.

The multiview image generation simulation system allows for many people to view the 3D image at the same time. The most basic way of creating multiview

images is to setup an array of cameras at each point-of-view and to take the picture at the same time [1]. These images are then processed to place them geometrically on the multiview display so that viewers can see a 3D scene at many different view-points. However, a number of multiview scenes captured by multiple cameras synchronously add more overhead and complexity. Furthermore, each camera needs to be adjusted because the intrinsic and extrinsic parameters of the cameras are all different.

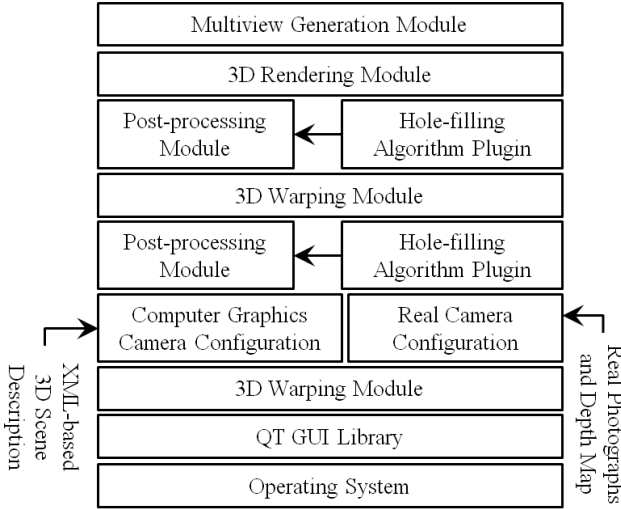
For these reasons, depth image based rendering (DIBR) is introduced as a way to generate multiple intermediate view images that look like they have been captured at various points of view [1]. In DIBR, the 2D color image with depth map is used together to synthesize a number of device-independent “virtual” views with different view angles and screen sizes of a scene, called as intermediate views. DIBR involved these three steps: pre-processing of depth map, 3D warping and creating multiview intermediate images and post-processing hole-filling. In prior works, a lot of pre- and post-processing on color images or depth maps are investigated to enhance the quality of the final multiview intermediate images [2, 3].

In this research, we developed a novel multiview display simulator system using the DIBR (Depth Image-Based Rendering) technique. This system supports generating multiview images of both computer graphics and real world scenes. It also supports options to plug in default or user-defined DIBR pre- or post-processing components. In DIBR, pre-processing often modifies the acquired depth map to increase the quality of the final intermediate images. Post-processing refers to the procedure for addressing the problem of occlusion areas by filling them with adjacent color information [4, 5]. In this paper, this system is used to evaluate two pre-processing depth map quantization methods.

In this paper, we will first describe the system overview of our DIBR-based multiview intermediate image generation simulator and then explain the design of linear and non-linear depth quantization method. We will then evaluate the quality of DIBR-based multiview images generated by using two depth map quantization methods. We will end with our conclusions and discuss directions for further research.

## 2 Multiview Image Generation Simulation System

Fig. 1 shows the overall architecture of multiview intermediate image generation system using the DIBR technique. It is built with Qt GUI and OSG 3D graphics library. It provides an XML-based script that allows constructing a 3D scene dynamically. When real photographs are used, both intrinsic and extrinsic parameters of the real color and depth cameras are used to generate images. Unlike previous works that only supported a fixed screen size, this simulator can be dynamically configured to suit various screen resolutions. This mechanism allows our system to be used from a mobile device to FULL HD TVs. In addition, this system can evaluate the effects of the quantization levels of a depth map image on the quality of the generated intermediate view images.



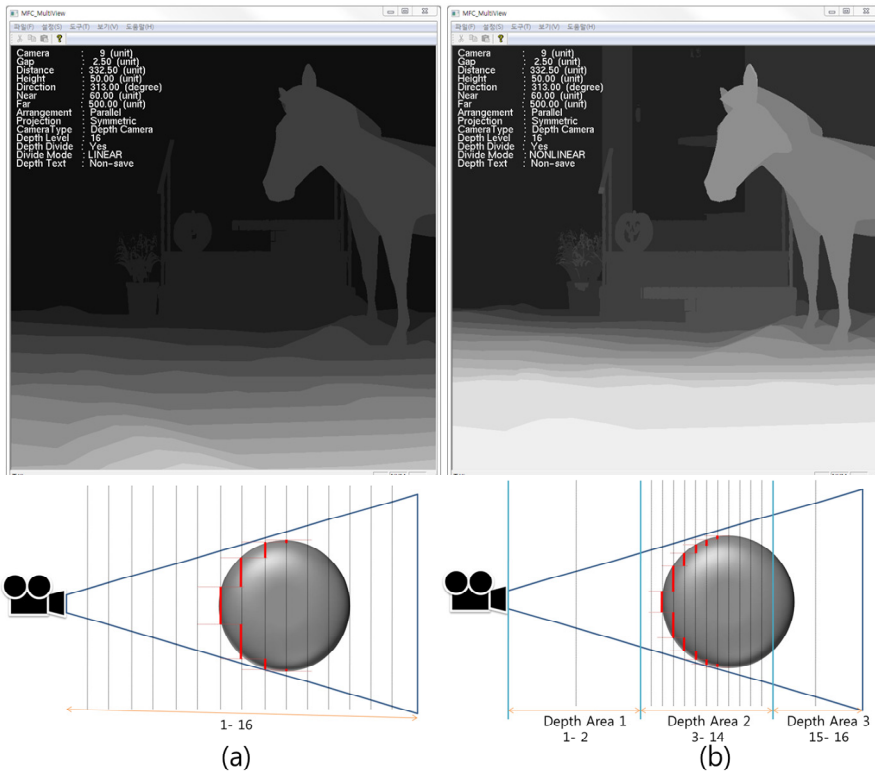
**Fig. 1.** The system architecture of multiview image generation simulation

In a computer graphics camera module, the 3D scene in XML format is rendered on the screen and the color and depth information are obtained from the scene. The camera module has a component for setting up the color and depth information and the camera’s intrinsic/extrinsic parameters. In this simulator, pre- and post-processing algorithms can be applied by inserting a simple plug-in. As shown in Fig. 1, this system simulates the entire series of DIBR steps, i.e. the input of a color/depth image, pre-processing of depth map, 3D warping to create a 3D point cloud image, and the use of the multiple virtual cameras to create multiview intermediate images. It then applies post-processing algorithms (i.e. hole-filling) to obtain the final multiview images.

Multiview image are usually composed of N-Views ( $N =$  a natural number greater than 1). Our simulation allows developers to easily adjust the number of virtual cameras that fits their needs. Furthermore, it is also possible to apply the camera’s intrinsic/extrinsic parameters to the virtual camera to simulate a scene that is very similar to one taken using the actual multiview cameras. Unlike other systems, our simulation system allows developers to freely adjust the resolution to fit the properties of the camera and multiview image display. In other words, it is possible to view the image on everything from a mobile device to a large display. The camera’s resolution can also be adjusted for simulation.

### 3 Linear and Non-linear Depth Quantization

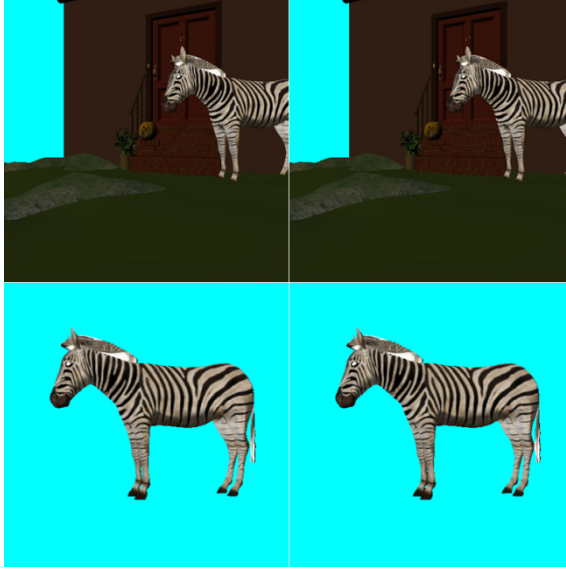
The accurate acquisition of depth image plays an important role in the DIBR process since it affects the quality of the restored 3D scene. If high quality depth image is



**Fig. 2.** Actual depth image and conceptual drawing of (a) linear (b) non-linear 4-bit depth map quantization (16 depth levels)

used, DIBR would gain high quality view images, but it also increases cost and complexity. It is found that about six to seven bit depths (64 to 128 depth levels) would be adequate enough in DIBR [6]. Higher than 7-bit depth would still increase the quality but the difference is only marginal. Linear depth map quantization simply divides depth levels evenly. However, as shown in Fig.2 (a), linear depth quantization may lose detail structures of the 3D object in the scene due to low bit depths. For instance, the object in Fig.2 (a) is laid in six out of sixteen depth levels; ten divisions are not effectively used.

This simulator also supports non-linear depth map quantization which enables to put more depth levels to the region of interest while sacrificing other areas. Fig.2 (b) shows the non-linear depth quantization that allocates fourteen levels out of sixteen to the “Depth Area 2,” which covers most portion of the object. When the depth image is used to restore the scene, non-linear depth quantization would certainly generate better quality than evenly distributed linear depth map quantization. Fig.2 (a) and (b) compares the actual depth image and the conceptual drawing of depth map divided by both linear and non-linear method. Fig.2 (b) shows more depths densely assigned in horse, doorsteps, and ground regions of the depth levels.

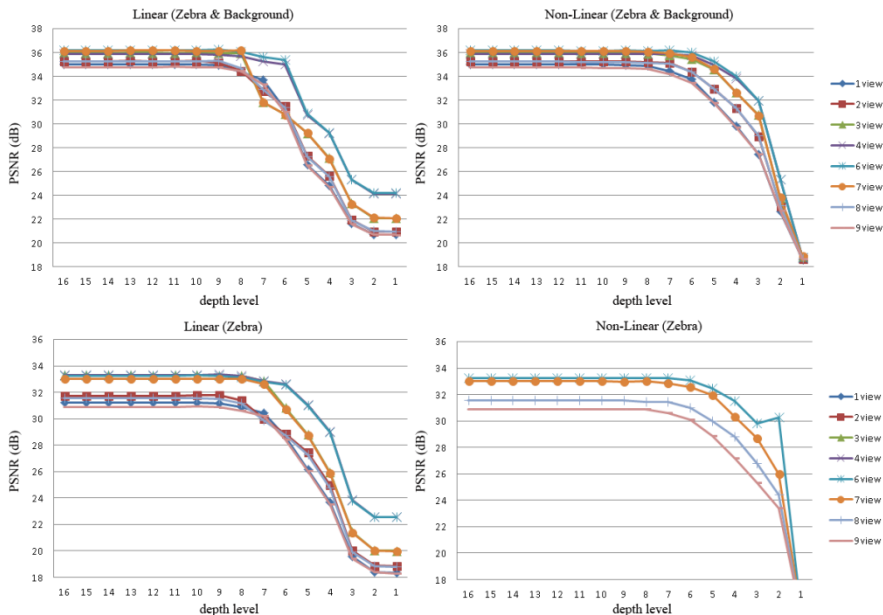


**Fig. 3.** Original image (left) and DIBR-based intermediate image (right)

Fig. 3 shows two 3D graphics scenes (Zebra and Zebra-object scene) of an original image directly captured by a virtual camera (left) and the DIBR-based multiview intermediate image (right) using 8-bit linear depth map quantization at view 7. An array of 9 cameras was placed in parallel at 3.25 unit intervals. The simulator obtained the color image and depth map from the fifth camera (center-view) and constructed the rest eight view intermediate images. The depth range,  $Z$  near and far bounds of the frustum, was fixed to 1000. The 16-bit to 1-bit depth map were applied for each of the 20 scenes. The depth quantization method, i.e. linear versus non-linear, was the independent variable, and the same hole-filling algorithm was used in this experiment.

Linear depth quantization took symmetric  $z$  value of depth buffer. On the other hand, non-linear depth quantization assigned more depth values (about 70 percent of depth precision) to the depth region where the Zebra object is located. Fig. 4 shows a comparison result of PSNR between the original images and the DIBR 9-view intermediate images by 16 depth quantization levels for linear (left) and non-linear (right) depth map quantization methods. The results revealed that there was no significant difference between two depth quantization methods when using 8-bit or higher depth level. On the other hand, non-linear depth quantization produced better performance at 7- to 3-bit depth level. However, non-linear depth quantization at lower than 3-bit depth level had led to worse results.





**Fig. 4.** PSNR between the original 9-view images and DIBR 9-view intermediate images by 16 quantization levels for two quantization methods

## 4 Conclusions

This paper presented a DIBR-based multiview intermediate image generation system. This system supports the generation of multiview intermediate images using either actual photographs or computer graphics scenes, making it that much more useful. Furthermore, it supports simple plug-in of pre-processing technique on depth image or post-processing (hole-filling) algorithm for the evaluation. Thus, researchers can use this system as a platform to compare and analyze different pre- and post-processing algorithms. We gave an overview of this system design: pre-processing depth map, 3D warping, and post-processing module.

Using this DIBR-based multiview image generation system, we evaluated the depth map quantization on two 3D computer graphics scenes, Zebra with background models and Zebra object only. Both linear and non-linear depth quantization methods were applied on these scenes. Then, the PSNR value between the 9-view original images directly captured from the scene and the DIBR-based 9-view intermediate images was measured to compare these two methods. The results showed that PSNR was between 30 and 36 dB, which were reasonable but not overly high. Overall, non-linear depth quantization produced better performance at 7- to 3-bit depth level.

It was determined that the PSNR value was easily affected by the hole-filling algorithm, lighting, and the low resolution of the obtained depth image. In the future, we plan to use this system to conduct more diverse research and to combine analytical

methods, such as PSNR or SSIM. We will also look into using GPU (3D warping) programming techniques to make the DIBR process faster.

**Acknowledgments.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Ministry of Education, Science and Technology (MEST) (No. 2012-0001640).

## References

1. Fehn, C.: Depth-Image-Based Rendering, Compression and Transmission for a New Approach on 3D-TV. In: Proc. of SPIE Stereoscopic Display and Virtual Reality Systems XI, January 19-21, vol. 5291, pp. 93–104 (2004)
2. Azzari, L., Battisti, F., Gotchev, A.: Comparative Analysis of Occlusion-Filling Techniques in Depth Image-Based Rendering for 3D Videos. In: Proc. of the 3rd Workshop on Mobile Video Delivery, pp. 57–62 (2010)
3. Oh, K.J., Yea, S., Ho, Y.S.: Hole-Filling Method Using Depth Based In-Painting for View Synthesis in Free Viewpoint Television (FTV) and 3D Video. In: Proceedings of 27th Conference on Picture Coding Symposium, pp. 233–236 (2009); Nguyen, Q.H., Do, M.N., Patel, S.J.: Depth Image-Based Rendering with Low Resolution Depth. In: Proc. IEEE, the 2nd ICIP, pp. 553–556 (2009)
4. Zhang, L., Tam, W.J.: Stereoscopic Image Generation Based on Depth Images for 3D TV. *IEEE Trans. on Broadcasting* 51(2), 191–199 (2005)
5. Kim, Y.-J., Lee, S.H., Park, J.-I.: A High-Quality Occlusion Filling Method Using Image Inpainting. *Journal of Broadcast Engineering* 15(1), 3–13 (2010)
6. Kim, M., Cho, Y., Choo, H.-G., Kim, J., Park, K.S.: Effects of Depth Map Quantization for Computer-Generated Multiview Images using Depth Image-Based Rendering. *KSII Transactions of Internet and Information Systems* 5(11), 2175–2190 (2011)

# Authoring System Using Panoramas of Real World

Hee Jae Kim and Jong Weon Lee

Department of Digital Contents, Sejong University, Seoul, Korea  
bisolby@naver.com, jwlee@sejong.ac.kr

**Abstract.** A panorama is a wide-angle view of a real world. Panoramas provide users real world information as the component of map services. Recently researchers try to augment additional information on panoramas to extend the usefulness of panoramas. However, the existing researches and applications provide users inconsistent experience by augmenting information on a single panorama. To solve this inconsistency, we present an authoring system helping users create contents on panoramas. Users create contents by augmenting virtual information on panoramas using the authoring system that propagates virtual information augmented on one panorama to neighboring panoramas. The resulting contents provide users consistent viewing experiences. Users can experience the contents on their desktop or they can view the contents on a smartphone display at the locations near to the locations panoramas were captured.

**Keywords:** Panoramas, Authoring, Augmenting, Consistent Experience.

## 1 Introduction

A panorama is a wide-angle view of a real world (Fig. 1). Panoramas have been used as the component of a map service such as Google Street View [1] and Microsoft Bing Maps Streetside [2] from the end of 2000 (Fig. 2). Panoramas provide users 360 degree views of a real world so users can understand the real environment of locations selected on a map. The coverage of panoramas extends to the world, even for off-roads and inside stores. Recently researchers try to augment additional information on panoramas to extend the usefulness of these panoramas. The Streetside photos of Microsoft [3] augments photos from Flickr on the Streetside view (Fig. 3). The photos are viewed as the part of the Streetdis view and provide users new experience. However, the existing researches and applications only augment additional information on a single panorama. The additional information is not presented on neighboring panoramas. This problem causes inconsistency in users' viewing experiences.

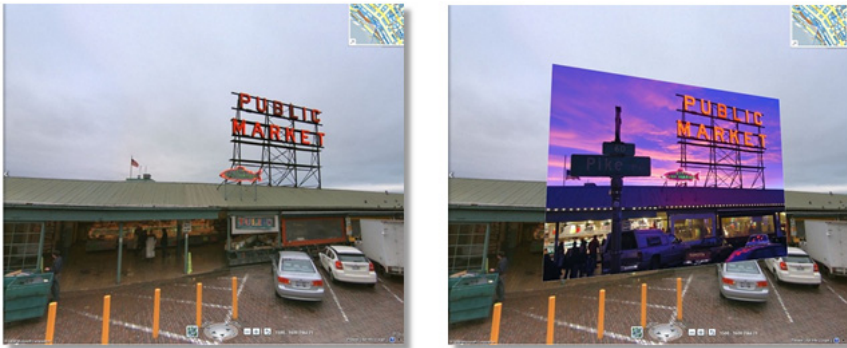
In this paper, we present an authoring system helping users create consistent contents on panoramas and share them with other users. Users create contents by augmenting virtual information on panoramas using the authoring system that propagates virtual information augmented on one panorama to neighboring panoramas. The resulting contents provide users consistent viewing experiences. Users can experience the contents on their desktop or they can view the contents on a smartphone display at the locations near to the locations panoramas were captured.



**Fig. 1.** The panorama of the old palace, Kyongbok Gung, in Korea

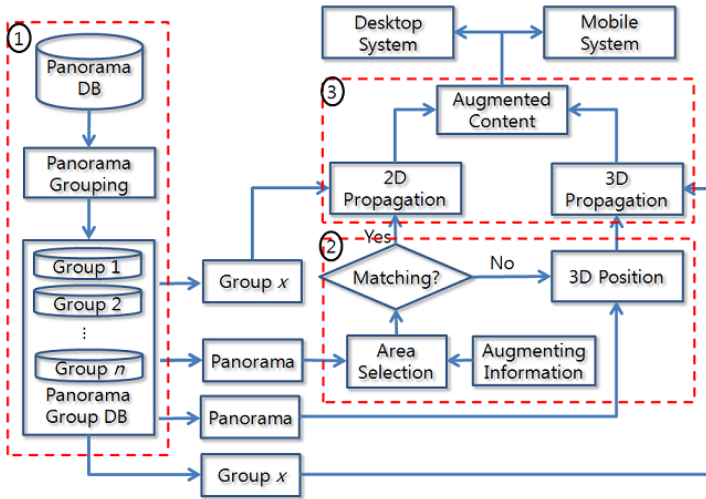


**Fig. 2.** Map services with panoramas (Left) Google Street View [1] (Right) Microsoft Bing Maps Streetside [2]



**Fig. 3.** Microsoft Streetside photos [3] (Left) Streetside view (Right) Streetside view with an augmented photo

## 2 Proposed System



**Fig. 4.** System overview (1) a grouping procedure (2) an area selection procedure (3) a propagation procedure

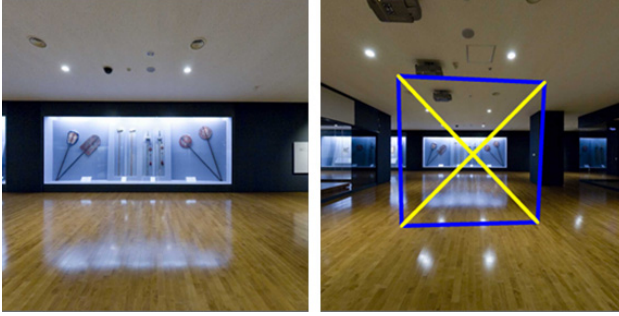
### 2.1 System Overview

The proposed system is divided into three procedures, a grouping procedure, an area selection procedure and a propagation procedure (Fig. 4). Panoramas are captured at indoor or outdoor environments and stored in a database (DB) then categorized into groups. A user selects panoramas and areas on the panoramas to augment additional information such as an image, a text, and a 3D object using the proposed system. The proposed system augments the additional information on the selected areas of the selected panoramas then propagates that information to other panoramas captured at locations near to the captured locations of the selected panoramas. Using this propagation procedure, the proposed system can help a user to create contents that provide consistent viewing experiences to users.

### 2.2 Grouping Procedure

After panoramas are stored in a DB, the grouping procedure groups panoramas in the DB into several groups and creates a panorama group DB. A user selects a key panorama from the DB. Panoramas in the DB captured at close locations to the key panorama’s location are grouped together and called a position-group. Panoramas in the position-group share scenes with the key panorama are grouped and called a sharing-group (Fig. 5). The matching algorithm SURF [4] is used to compare panoramas in the position-group with a key panorama. The performance of matching between panoramas is lower than the performance of matching between perspective images because of the deformation on panoramas. Panoramas are transformed into four perspective

images before applying the matching algorithm. These position-groups and sharing-groups are input to the area selection and the propagation procedures. The grouping procedure runs once so it does not cause any delay to the authoring systems.



**Fig. 5.** Creating a sharing-group (Left) one of four perspective images of the key panorama (Right) one of four perspective images of the matched panorama with the rectangle indicating the view of the left image

### 2.3 Area Selection Procedure

A user selects an area on a panorama (called PA) to augment virtual information in the area selection procedure. The user selects the target area by browsing the panorama PA with the panorama viewing system, which displays the perspective view of the panorama. If the panorama PA is found in the sharing-groups, the target area is called 2D augmenting area and the area selection procedure is ended. If the panorama PA is found in the position-groups only, the procedure asks the user to select the same target area on another panorama (called PB). The 3D positions of the target area are computed by finding intersections of two pair of lines.

Panoramas were captured with their positions and orientations. The azimuths and altitudes of panoramas are used to align all panoramas to the predefined direction using the equation 1. The 3D positions are computed using two angles ( $A_{pc}$ ,  $a_{pc}$ ) and locations of two panoramas. The area defined by the estimated 3D positions is called a 3D augmenting area (indicated as two circles in the top and the middle images in Fig. 6).

$$\begin{cases} A_p + I_h \frac{360}{h} = A_{pc} \\ a_p + \frac{180}{v} \left( I_v - \frac{v}{2} \right) = a_{pc} \end{cases} \quad (1)$$

$h$  and  $v$  indicate the size of a panorama in pixels along the horizontal and vertical direction respectively,  $(I_h, I_v)$  indicates the location of a target pixel,  $A_p$  and  $a_p$  are azimuth and altitude of the camera, and  $A_{pc}$  and  $a_{pc}$  are compensated azimuth and altitude of the target pixel.



**Fig. 6.** An image augmented on a 3D target area (Top) a panorama PA and one perspective view of PA with two selected points (blue circles) (Middle) a panorama PB and one perspective view of PB with two selected points (blue circles) (Bottom) The 3D target area (a red rectangle) on another panorama and one perspective view showing an augmented content

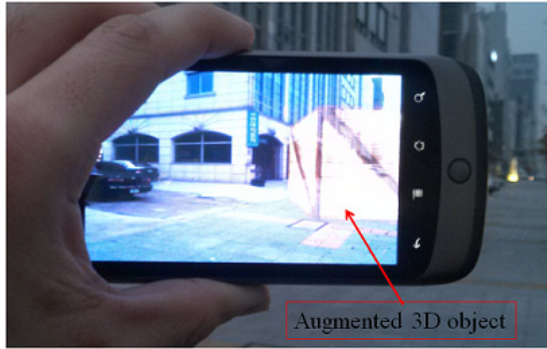
## 2.4 Propagation Procedure

Augmenting areas and corresponding contents are propagated to other panoramas in the group DB using the propagation procedure. It is divided into 2D propagation and 3D propagation procedures. In the 2D propagation procedure, 2D augmented area is searched on panoramas in the sharing-group using the matching algorithm. If the corresponding area is found, the content is augmented on the detected area. In the 3D propagation procedure, the image location of the 3D augmenting area is estimated using the azimuth, altitude and the position of each panorama and positions of the 3D augmenting area. The azimuth and altitude of the panorama is used to align the panorama to the same predefined direction.

## 2.5 Viewing Contents

The augmented contents on panoramas can be viewed on the desktop environment and the mobile environment. In the desktop environment, a user will browser each panorama one by one and the user can view the content on consecutive panorama not like the exiting system providing a user a single view of the augmented content. One example view is shown in Fig. 6. In the mobile environment, a user views the augmented content on the real camera view of the mobile phone (Fig. 7). The content is augmented based on the position and the azimuth and the altitude of the mobile phone and the positions of 3D augmenting area.

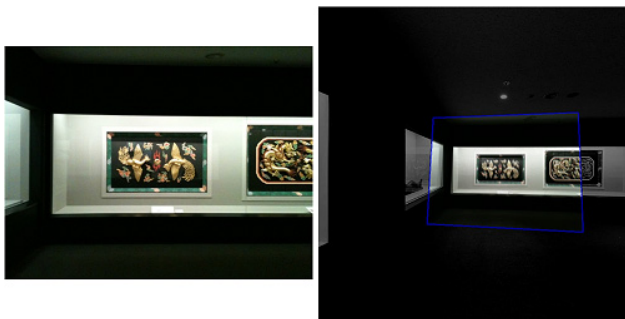




**Fig. 7.** Viewing an augmented 3D object on a mobile phone

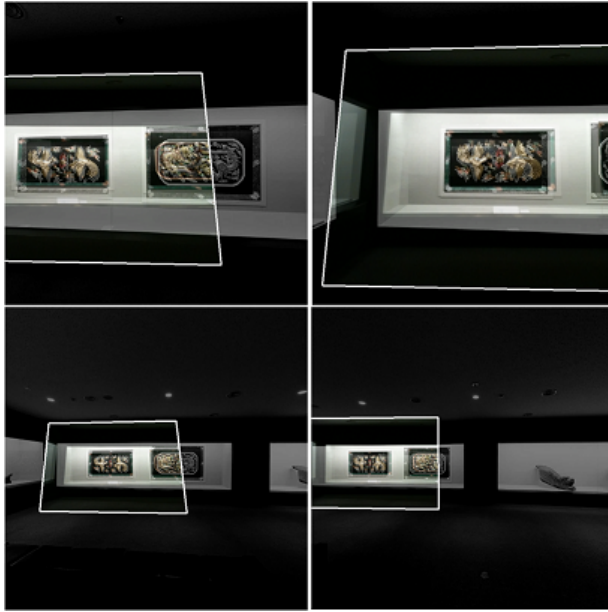
### 3 Experiments

We applied the proposed system in three locations, one indoor and outdoor locations. First the proposed system was used to propagate a target area to panorama in the sharing group. Since the matching between images did not work well for images captured at outdoor environment, we used panoramas captured at the museum to test the 2D propagation procedure. The distances between two consecutive panoramas captured at the museum were about three meters. A user selected a part of an image as the target area, which was used to augment a virtual object. The user also selected the first target panorama and confirmed the correct augmentation on the selected panorama. The augmentation result on the selected panorama is shown in Fig. 8. If the augmentation was not correct, the user could modify the location of the augmented object using a mouse. Using the 2D propagation procedure, the 2D augmenting areas on other panoramas were detected and the virtual object was augmented on the 2D augmenting areas (Fig. 9). This result demonstrated that the proposed system could easily propagate augmented contents to other panoramas using the 2D propagation procedure of the proposed system.



**Fig. 8.** An experiment at the museum (Left) a target area (Right) augmenting a quadrilateral on the 2D augmenting area found on the target panorama





**Fig. 9.** The results of 2D propagation with white quadrilateral, which are augmented virtual information

The proposed system was applied to panoramas captured at open space in an old palace. The distances between two consecutive panoramas captured at the old palace were about five meters and the result is shown in Fig. 6. A user selected the same target area on two panoramas and augmented the virtual object on the target areas (Fig. 6). The augmented object was propagated to other panoramas using the 3D propagation procedure of the proposed system (Fig. 6). One example view on a mobile device is shown in Fig. 7. This experiment demonstrated the proposed system could be used to augment virtual objects on panoramas easily and so provide the user consecutive viewing experience.

## 4 Conclusion

The proposed system was tested with panoramas captured in indoor and outdoor environments. The 2D propagation was used to augment contents on panoramas captured inside the museum because the matching between panoramas was quite successful. Since the matching between panoramas captured in the outdoor environment was poor, the 3D propagation was used to augment contents on panoramas captured at an old palace. The augmented content was viewed on a PC and a smartphone with GPS and the rotation sensor for outdoor experiment.

The proposed system helped users create useful contents on panoramas and provided consistent viewing experience. The proposed system also had few limitations. The first limitation is the accuracy of GPS. GPS was used to estimate positions of

panoramas that are used to estimate positions of 3D augmenting areas. Because of the poor estimation of the positions of 3D augmenting area, the augmented results are sometimes not realistic. We need to overcome this limitation using other information on the images since the accuracy of GPS is not going to improve soon. Another limitation is the accuracy of the matching algorithm. Currently the matching algorithm is not applicable for outdoor environment and some indoor environment. If the accuracy of the matching algorithm is improved, we can use the matching algorithm more frequently in the proposed system to create contents.

**Acknowledgements.** This work (Grants No. C0016661) was supported by Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2012. This work was also supported by the Industrial Strategic technology development program, 10041772, (The Development of an Adaptive Mixed-Reality Space based on Interactive Architecture) funded by the Ministry of Knowledge Economy (MKE, Korea).

## References

1. Google Maps Street View, <http://maps.google.com>
2. Microsoft Bing Maps Streetside, <http://www.microsoft.com/maps/streetside.aspx>
3. Streetside Photos of Microsoft Bing Maps, [http://www.bing.com/community/site\\_blogs/b/maps/archive/2010/02/11/new-bing-maps-application-streetside-photos.aspx](http://www.bing.com/community/site_blogs/b/maps/archive/2010/02/11/new-bing-maps-application-streetside-photos.aspx)
4. Bay, H., Ess, A., Tuytelaars, T.: L. Gool, L.: Speeded-up robust features (surf). *Journal of Computer Vision* 110(3), 346–359 (2008)

# Integrated Platform for an Augmented Environment with Heterogeneous Multimodal Displays

Jaedong Lee, Sangyong Lee, and Gerard Jounghyun Kim

Digital Experience Laboratory  
Korea University, Seoul, Korea  
{jdlee, xyleez, gjkim}@korea.ac.kr

**Abstract.** With the recent advances and ubiquity of various display systems, one may configure an augmented space with a variety of display systems, such as 3D monitors, projectors, mobile devices, holographic displays, and even non-visual displays such as speakers and haptic devices. In this paper, we present a software support platform for representing and executing a dynamic augmented 3D scene with heterogeneous display systems. We extend the conventional scene graph so that a variety of modal display rendering (aside from just visual projection) can be supported. The execution environment supports multi-threading of the rendering processes for the multiple display systems and their synchronization. As multiple and heterogeneous displays, in effect representing a particular set of objects in the augmented environment, are scattered in the environment, an additional perception based spatial calibration method is also proposed.

**Keywords:** Augmented space, Extended scene graph, Multiple displays, Calibration, Floating image display.

## 1 Introduction

An augmented environment refers to a physical 3D environment spatially and naturally registered with virtual objects. Being “natural” means that the virtual objects are perceived to mix in with the physical environment seamlessly and felt as everyday objects. Tight spatial registration means virtual objects situated in the right location and pose, and this would be one requirement for naturalness. With the recent advances and ubiquity of various display systems, realization of such “naturally” augmented environments has become viable. One may configure an augmented space with a variety of display systems, such as 3D monitors, projectors, mobile devices, holographic displays, and even non-visual displays such as speakers and haptic devices (Figure 1). In this paper, we present a software support platform for representing and executing a dynamic augmented 3D scene with heterogeneous display systems. We extend the conventional scene graph so that a variety of modal display rendering (aside from just visual projection) can be supported. The execution environment supports multi-threading of the rendering processes for the multiple display systems and their synchronization. As multiple and heterogeneous displays scattered in the environment does not easily lend themselves to the conventional, e.g. computer vision, based calibration process, a perception based spatial calibration method is proposed.



Fig. 1. An augmented environment with multiple heterogeneous displays

## 2 Related Work

Virtual reality environments often require their presentations through large displays. For that purpose, researchers have proposed methods for distributed rendering and tiled display [1, 2, 3]. However, these systems typically do not address the use of different types of display systems, nor do they consider displays systems that are dispersed in the environment through which augmentation environment/object can be viewed. One notable exception is the work by Yang et al. who proposed a layered display system [4] and switching between heterogeneous display systems in the environment as the users moves around in it for an optimized viewing condition. In our work, we focus on a comprehensive (in terms of types of display systems supported, synchronization, and calibration) software platform for handling multiple heterogeneous displays.

## 3 Scene Graph Extension

### 3.1 Object Types / Display Parameters

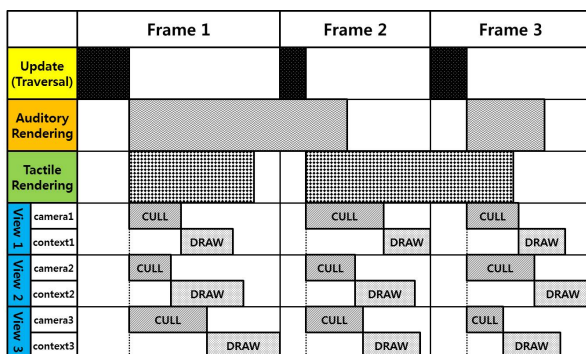
We took a conventional scene graph data structure, and extended it so that it can represent different types of objects and specify associated display systems and required parameters. For instance, an object type may be a real physical object with its geometry and attributes captured by sensors, reconstruction methods and even direct measurements. An object may be purely virtual (augmented into the environment) and designated to be presented visually through a holographic device and with tactile feedback through a vibration device. Note that for a given object and its modality, the developer may wish to specify a particular display rendering algorithm. Node types and attributes have been revised support specification of such information and parameters so that when the scene graph is processed, it can be used for proper

rendering and synchronization. With the extended node sets, the developer can specify and design the augmented environment with more ease and without having to worry about low level details.

### 3.2 Synchronization

There are two main objectives for synchronization. One is for among image frames of different visual displays, and the other for among different modality output corresponding to a particular event. For example, when virtual ball is dropped, it may be rendered visually, aurally and with force feedback, and all the modal output must occur with minimal temporal delay to be felt as one event.

For the former, similarly to the techniques employed by tiled display systems, we use software “gen-lock” to synchronize image frames among within 2~3 frame difference [5, 6]. Such degree of difference is usually regarded visually unnoticeable. In our case, the slowest rendering node serves as the reference to all other rendering nodes for synchronization (see Figure 2). While frame coherence is guaranteed, the temporal coherence may suffer if there were many rendering nodes to which network messages must be sent. However, in normal situations, there would not be so many rendering nodes (e.g. less than 20~ 30). What is important is that upon an event its multimodal after effects are rendered simultaneously. When different visual displays are rendered on separate graphics card, the hardware gen-lock can be used [7].

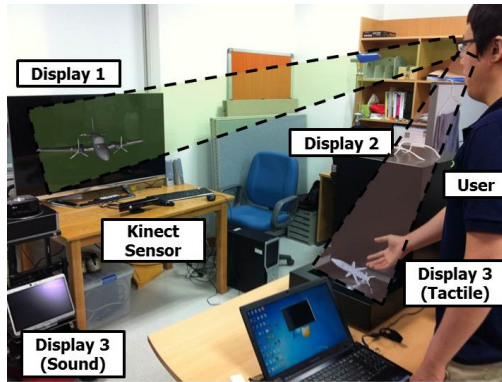


**Fig. 2.** Three rendering nodes synchronized. For example, the visual rendering of Frame 2 only starts after view 3 has finished rendering Frame 1. In the meantime, the aural and tactile rendering starts off upon traversing the scene graph and invoking the corresponding call back threads, which may sometimes continue over the visual frame boundary. The aural and tactile rendering, however, does not occur as frequently as the visual (which must occur at least 15~20 times per second).

As for the latter objective, our extension provides a protocol for different modal displays to refer to a common event specification and synchronize each multimodal output thread around it by a call-back mechanism. For example, the aural and tactile rendering starts off after traversing the scene graph by restarting the corresponding the call-back threads, which may sometimes continue over the visual frame boundary.

Practically, this is not problematic as the aural and tactile events and their rendering, however, do not occur as frequently as the visual (which must occur at least 15~20 times per second).

A test system using the proposed platform has been set up and has been evaluated in terms of the overall frame rate (vs. number of displays sustainable) and temporal synchronization error. Figure 3 shows the test system in which three (and more) different displays are used, a 3D stereoscopic TV (showing the airplane), floating image device (showing the missile) and vibro-tactile device. The objects are spatially registered and synchronized such that when interacted upon, the missile would render a tactile feedback, fly and shoot down the airplane. Our have shown acceptable performance (up to ~20 fps) for supporting up to 12 different displays with each node handling up to more than 120,000 polygons, and exhibiting inter-node frame coherence, and virtually no temporal delay. Such a performance level is deemed sufficient to support and implement a small augmented room.



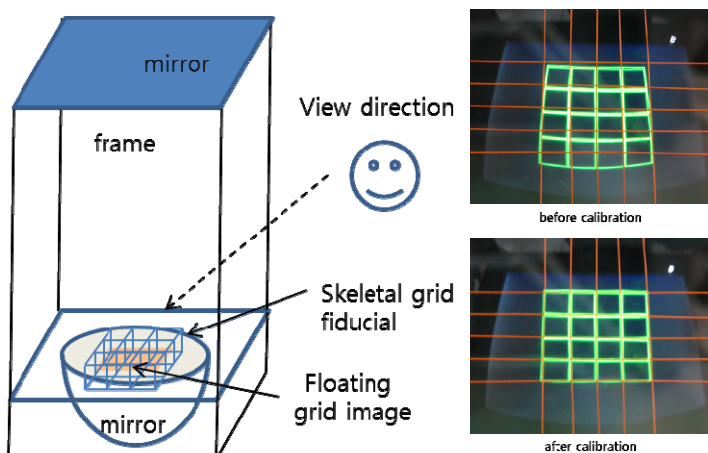
**Fig. 3.** Test environment consisting of three (and more) different displays (3D TV, floating image display, vibro-tactile device, etc.) in an augmented environment.

## 4 Calibration

As virtual objects are situated and registered to the physical world through associated display systems, there needs to be a method for these display system to be calibrated according to the units of the physical world and spatially registered in the whole environment. While standard calibration methods and warping methods exist for monitors and projectors, since the augmented space also use other types of displays, e.g. holographic floating image device, we devised a new calibration method, since its display mechanism is different from that of the monitors/projectors. The main difference is that its optical system floats an image in 3D space and it is very difficult to establish the correspondence between the ground truth and actual display points due to the geometry of the display apparatus and other restrictions. For example, the floating image cannot be seen from the front direction due to the location of the reflecting mirrors and other optical elements. The ground truth object has to be

suspended in 3D space in a “skeletal” form not to occlude any reflecting lights in forming the floating image (see Figure 4). Installation of such a grid structure object is not always possible. Then, one could take a picture, from a known view point location, of the floating image overlaid on the ground truth skeletal object, and apply image processing techniques to extract the “points” and a set of correspondence match. However, again this is problematic because the imagery is usually not conducive to corner point extraction (e.g. the interior of the display device is dark and the skeletal grid points are barely visible).

Instead we rely on a human to judge whether the floating grid points coincide with those of the ground truth grid. Any perceptual differences are made coincident by adjusting the corresponding points in the virtual space. Such differences (between the virtual points of the ground truth and the virtual points of the adjusted ones) are recorded and later applied for image correction by interpolating the adjustment values. Figure 4 also shows the results of applying the calibration. Note that such calibration is inevitably view dependent and user dependent because the optical system for the floating imagery is complicated resulting in different distortion depending on the view point and because we rely on human judgments (particularly in depth assessments). Thus, ideally, the calibration must be performed at different nominal view points and for customized to the individual user.



**Fig. 4.** Calibrating the floating image device used in our augmented space. A skeletal grid representing the ground truth is attached and suspended in the middle of the display space above the spherical mirror. A floating grid image is compared to the ground truth and made coincident perceptually from a given view point by adjusting the corresponding points in the virtual space. The adjustments are recorded and applied to other objects through interpolation.

## 5 Conclusion

In this paper, we have presented a software support platform for representing and executing a dynamic augmented 3D scene with heterogeneous display systems.

While more performance testing and optimization is required, it offers a convenient software layer abstraction for realizing augmented environments. We hope that such an infrastructure will contribute to proliferate the use of augmented environments for various applications to such as entertainment and education.

**Acknowledgement.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No 2012-0009232)

## References

1. Li, C., Lin, H., Shi, J.: A survey of multi-projector tiled display wall construction. In: Multi-Agent Security and Survivability, pp. 452–455 (2004)
2. Ni, T., Schmidt, G.S., Staadt, O.G., Livingston, M.A., Ball, R., May, R.: A survey of large high-resolution display technologies, techniques, and applications. In: Virtual Reality Conference, pp. 223–236 (2006)
3. Krumbholz, C., Leigh, J., Johnson, A., Renambot, L., Kooima, R.: Lambda table: high resolution tiled display table for interacting with large visualizations. In: 5th Workshop on Advanced Collaborative Environments (2005)
4. Yang, U.Y., Kim, H.M., Kim, J.H.: Expandable 3D Stereoscopic Display System. Korea Patent Applied, No. 2012-0018690 (2012)
5. Van Der Schaaf, T., Renambot, L., Germans, D., Spoelder, H., Bal, H.: Retained mode parallel rendering for scalable tiled displays. In: Immersive Projection Technologies Symposium (2002)
6. Igehy, H., Stoll, G., Hanrahan, P.: The design of a parallel graphics interface. In: 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 141–150. ACM (1998)
7. Nvidia, Inc. Genlock, [http://www.nvidia.com/object/IO\\_10793.html](http://www.nvidia.com/object/IO_10793.html)



# Optimal Design of a Haptic Device for a Particular Task in a Virtual Environment

Jose San Martin<sup>1</sup>, Loic Corenthy<sup>2</sup>, Luis Pastor<sup>1</sup>, and Marcos Garcia<sup>1</sup>

<sup>1</sup>Universidad Rey Juan Carlos, Madrid

<sup>2</sup>Universidad Politecnica, Madrid  
jose.sanmartin@urjc.es

**Abstract.** When we create an environment of virtual reality based training that integrates one or several haptic devices sometimes the first choice to make is the device to use. This paper introduces an algorithm that allows us, for a particular task to be simulated in a virtual environment, to find key data for the design of appropriate haptic device, or to select the clues in order to get optimum performance for that environment and that particular task.

**Keywords:** Virtual Reality, Haptics workspace, Manipulability, Optimal designing.

## 1 Introduction

Learning based on virtual reality (VR) is widespread in the field of training of different techniques, such as surgery [1-6]. In this field significant improvements have been obtained in the combination of traditional learning with VR based simulators [7, 8]. The use of this type of simulators has spread to other techniques [9, 10] apart from entertainment [11].

When finding the haptic device which is more suitable for our task we consider two possible ways: On the one hand we have the ability to design a custom haptic device, so that the system requirements are the conditions of the design of the new device [12] or may be composed of various devices [13]. Moreover we will be unable to create a new device but we need a tool that allows us to choose between different haptic devices, one that best suits our needs.

The question to answer is whether we can nevertheless find the haptic device suitable for our design. The first thing we have to study is the virtual environment in which we work. If you want a versatile system that includes many different environments, we must choose a device obviously generalist, but it is possible that this device is pretty good at all, but not the best in any of the environments.

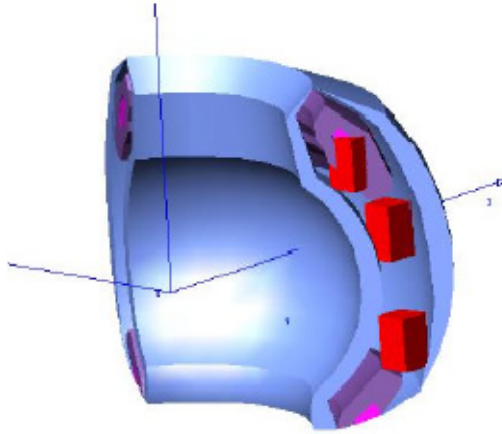
There are several methods to optimize the design of a manipulator and to evaluate the suitability a haptic device in a specific application [14]. Frequently, a criterion consists of obtaining the highest Manipulability measure in the whole workspace [15-20]. The contribution includes several measures of quality of the mechanical design of a virtual training system [21, 13].

In this paper we present an algorithm that allows an easy measurement of the proper dimensions of a manipulator-type device, to work on a particular task.

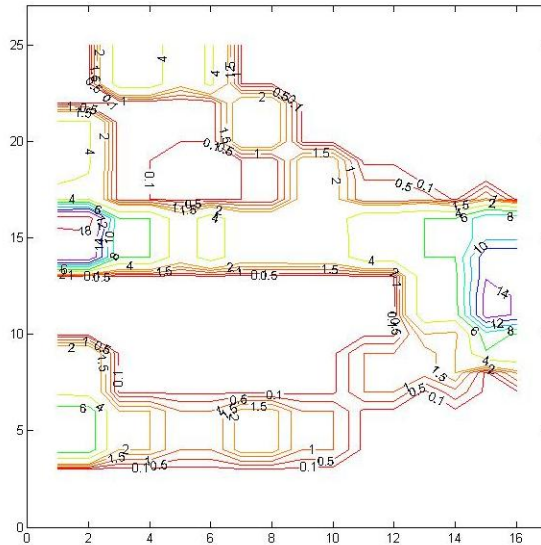
Finally we present the results in each case allowing the key data in order to design the optimal haptic device for each duty.

## 2 Defining the Virtual Environment

A task to be performed in a simulated environment can be defined by two characteristics: first the virtual environment (VW), the volume where the simulation is performed (Fig. 1), the space in which the End Effector (EE) is moving.



**Fig. 1.** Detail of positioning options of a VW inside the RW



**Fig. 2.** Section 2D of the NFM corresponding with the simulation of the working of a machine tool

Moreover, the movement within this environment is not homogeneous and define a navigation frequency map (NFM) according to the zones of VW the EE is visiting (fig. 2).

Figure 1 shows the problem to be solved: first we have the volume (RW) represented by the points in space that can reach a haptic device with EE. Of course the VW is smaller than the RW. We must find out what part of RW we select to work, given that the distribution in RW is not homogeneous.

### 3 Study of the Quality of the Workspace

The VW is a subset of all the space a haptic device can achieve: the Real Workspace (RW). This space is a characteristic of each device, and depends on their mechanical properties. Because of this, the space is not homogeneous, for instance, near the singular points of the mechanism, we find points where the quality of the device efficiency is very low.

To quantify the efficiency of the device, we will implement different measures based on the concept of Manipulability [22][23][24]:

$$\mu = \sigma_{\min}(J_u)/\sigma_{\max}(J_u) \quad (1)$$

where  $\sigma_{\min}(J_u)$  and  $\sigma_{\max}(J_u)$  are the minimum and maximum singular values of  $J_u$ .  $\mu \in [0;1]$  being 1 the optimal value.  
or the average value

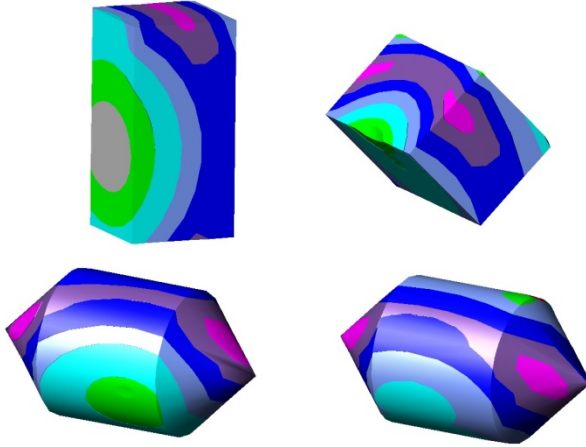
$$\mu_v = \frac{\int_0^{V_T} \mu_i \cdot v_i \cdot dv}{V_T} \quad (2)$$

where  $\mu_i$  is the value of Manipulability in each sub-volume  $v_i$  of iso-Manipulability

### 4 Problem to Be Solved

Since VW is a subset of RW can also study the problem in reverse, that is, defining the minimum quality required in each of the proposed virtual environments (fig.3-1, 3-2). From that minimum desirable quality, and each VW, we can design the right device.

The desired design is based on a device similar constructively to PHANTOM family. First we draw the volume enclosing the VW, so any point of the virtual environment is not beyond the range of the device. More important is the quality of the workspace. To build that RW will be used spheres of uniform value of Manipulability.



**Fig. 3.** Two different examples of VW. In each VW we can see subsets in different colors representing different values of desired Manipulability

## 5 Methodology

Firstly it is necessary to define in the VW to study the average desired values, as shown in Figures 3. The proposed method begins from the ideal configuration of a manipulator, that is with both arms of equal value  $L$  (unknown already and defined in the algorithm) forming an angle of  $90^\circ$ . As shown in the map of Figure 4, Manipulability maximum values coincide with that configuration.

We place the center of the subset of VW where the best values are requested, exactly in the EE of the initial configuration described above. We will define the RW from this point. Around this point we create a sphere of initial radius 2 mm (in this space can be assumed constant value calculated by (2)) and the algorithm begins:

1° It is increased the size of the sphere (the initial resolution 1mm).

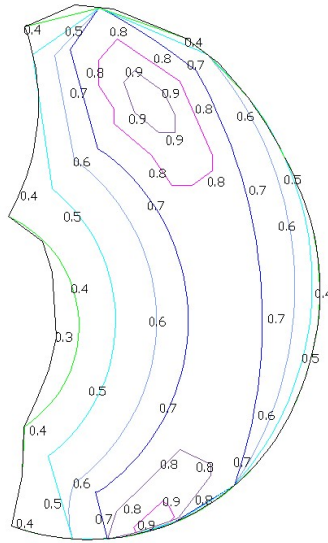
2° Mean value of Manipulability is checked in the current sphere.

2-1 If the average value is greater than or equal to the required value:

- It must be check that the sphere includes the all VW zone with the iteration value of Manipulability.
  - o Yes: Next subset of VW.
  - o No: the area is still too small, it is increasing the radius of the sphere. Step 1.

2-2 If the average value is lower than the required value:

- We modify the value  $L$ , the considered length of the arms of the manipulator and proceed to step1 recalculating the sphere of radius 2mm.



**Fig. 4.** Subspace 2D of Manipulability defined for the real workspace of an OMNi

After the first iteration and an initial value of  $L$ , is passed to the next subset of  $VW$  with the immediately lower Manipulability value. Once all subsets we can divide  $VW$  are studied, the algorithm terminates.

The object of the algorithm is twofold, first the manipulator, and  $RW$  should reach all points of  $VW$ , and the other, in each zone should be achieved with a minimum efficiency value.

## 6 Results

In order to evaluate the result of the algorithm, there are three different examples (Simulating arthroscopic surgery, simulation of a boiler inspection, and operation of a machine tool-figure 2) with different sizes of  $VW$  and different tasks within. We check that the haptic device designed for each job, has different mechanical characteristics.

Case1.- Simulating arthroscopic surgery. Best value of  $L=142$  mm. Similar to Sensable's PHANToM OMNi.

Case2.- Simulation of a boiler inspection. Best value of  $L= 118$  mm. Similar to Force Dimension's OMEGA.

Case3.- Operation of a machine tool. Best value of  $L= 129$  mm. Similar to Sensable's PHANToM OMNi.

## 7 Conclusions

It has presented an algorithm that allows the optimal design of a haptic device that will be used for a specific simulation task.

It has been determined that if this task varies, the most suitable device has different dimensions.

When confronted with the task of a simulation involving a haptic device, we can use the path defined in this paper as well to design the best possible device properly or to choose from a set of existing devices.

**Acknowledgements.** This work has been partially funded by the FP7 Integrated Project Wearhap: Wearable Haptics for Humans and Robots and the Cajal Blue Brain Project.

## References

1. Agha, R., Muir, G.: Does laparoscopic surgery spell the end of the open surgeon? *J. R. Soc. Med.* 96, 544–546 (2003)
2. Ahlberg, G., Heikkinen, T., Iselius, L., Leijonmarck, C.E., Rutqvist, J., Arvidsson, D.: Does training in a virtual reality simulator improve surgical performance? *Surg. Endosc.* 16, 126–129 (2002)
3. Rosen, J., Hannaford, B., MacFarlane, M.P., Sinanan, M.: Force controlled and teleoperated endoscopic grasper for minimally invasive surgery-experimental performance evaluation. *IEEE Transactions on Biomedical Engineering* 46 (1999)
4. Burdea, G., Patounakis, G., Popescu, V., Weiss, R.E.: Virtual reality-based training for the diagnosis of prostate cancer. *IEEE Transactions on Biomedical Engineering* 46 (1999)
5. Foderoli, K., King, H., Lum, M., Bland, C., Rosen, J., Sinanan, M., Hannaford, B.: Control system architecture for a minimally invasive surgical robot. In: *Proceedings of Medicine Meets Virtual Reality* (2006)
6. Ward, J., Wills, D., Sherman, K., Mohsen, A.: The development of an arthroscopic surgical simulator with haptic feedback. *Future Generation Computer Systems* 550, 1–9 (1998)
7. Grace, P., Borley, N., Grace, P.: *Surgery at a Glance*. Blackwell Science (2002)
8. Alfonso, C.D., Blanquer, I., Segrelles, D., Hernandez, V.: Simulación quirúrgica sobre escenarios realistas. In: *Proceedings of Congreso Nacional de Informática Médica-Informada* (2002)
9. Immonen, L.: *Haptics in Military Applications*. Diss. University of Tampere (December 2008); Web (November 8, 2012)
10. Jiang, L., Girotra, R., Cutkosky, M., Ullrich, C.: Reducing Error Rates with Low-Cost Haptic Feedback in Virtual Reality-Based Training Applications. In: *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. World Haptics (2005)
11. Fogg, B.J., Cutler, L.D., Arnold, P., Eisbach, C.: HandJive: a device for interpersonal haptic entertainment. In: *CHI 1998: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 57–64 (1998)
12. Müller, W., Bockholt, U., Lahmer, A., Voss, G., Börner, M.: VRATS Virtual Reality Arthroscopy Training Simulator. *Radiologe* 40, 290–294 (2000)
13. SanMartin, J., Trivino, G., Bayona, S.: Mechanical design of a minimal invasive surgery trainer using the manipulability as measure of optimization. In: *Proceedings of IEEE International Conference on Mechatronics, ICM* (2007)
14. Sobh, T., Toundykov, D.: Optimizing the tasks at hand robotic manipulators. *IEEE Robotics & Automation Magazine* 11(2), 78–85 (2004)

15. Alqasemi, R., McCaffrey, E., Edwards, K., Dubey, R.: Analysis, evaluation and development of wheelchair-mounted robotic arms. In: Proceedings of International Conference on Rehabilitation Robotics, ICORR (2005)
16. Guilamo, L., Kuffner, J., Nishiwaki, K., Kagami, S.: Manipulability optimization for trajectory generation. In: Proceedings of IEEE International Conference on Robotics and Automation, ICRA (2006)
17. Masuda, T., Fujiwara, M., Kato, N., Arai, T.: Mechanism configuration evaluation of a linear-actuated parallel mechanism using manipulability. In: Proceedings of IEEE International Conference on Robotics and Automation (2002)
18. Bayle, B., Fourquet, J.Y., Renaud, M.: Manipulability of wheeled mobile manipulators: Application to motion generation. *The International Journal of Robotics Research* 22, 565–581 (2003)
19. Liu, H., Huang, T., Zhao, X., Mei, J., Chetwynd, D.: Manipulability of wheeled mobile manipulators: Application to motion generation. *Mechanism and Machine Theory* 42, 1643–1652 (2007)
20. Wang, S., Yue, L., Li, Q., Ding, J.: Conceptual design and dimensional synthesis of “microhand”. *Mechanism and Machine Theory* 43, 1186–1197 (2008)
21. Martin, J.S.: A study of the attenuation in the properties of haptic devices at the limit of the workspace. In: Shumaker, R. (ed.) VMR 2009. LNCS, vol. 5622, pp. 375–384. Springer, Heidelberg (2009)
22. Yoshikawa, T.: *Foundations of Robotics: Analysis and Control*. MIT Press, Cambridge (1990)
23. Cavusoglu, M.C., Feygin, D., Tendick, F.: A critical study of the mechanical and electrical properties of the phantom haptic interface and improvements for high performance control. *Teleoperators and Virtual Environments* 11, 555–568 (2002)
24. Yokokohji, Y., Yoshikawa, T.: Guide of master arms considering operator dynamics. *Journal of Dynamic Systems, Measurement, and Control* 115(2A), 253–260 (1993)

# Real-Time Dynamic Lighting Control of an AR Model Based on a Data-Glove with Accelerometers and NI-DAQ

Alex Rodiera Clarens<sup>1</sup> and Isidro Navarro<sup>2</sup>

<sup>1</sup>Dpto. de Arquitectura, Universidad Ramon Llull, Barcelona, Spain  
alex.rodiera@gmail.com

<sup>2</sup>Dpto. de Expresión Gráfica Arquitectónica–Escuela Técnica Superior de  
Arquitectura de Barcelona, Universidad Politécnica de Cataluña, Barcelona, Spain  
isidro.navarro@upc.edu

**Abstract.** The lighting of models displayed in Augmented Reality (AR) is now one of the most studied techniques and is in constant development. Dynamic control of lighting by the user can improve the transmission of information displayed to enhance the understanding of the project or model presented. The project shows the development of a dynamic control of lighting based on a data-glove with accelerometers and A/D NI-DAQ converter. This device transmits (wired/wirelessly) the signal into the AR software simulating the keystrokes equivalent to lighting control commands of the model. The system shows how fast and easy it is to control the lighting of a model in real-time following user movements, generating great expectations of the transmission of information and dynamism in AR.

**Keywords:** Real-time lighting, NI-DAQ, Accelerometers, Xbee, Data-glove, augmented reality.

## 1 Introduction

Augmented Reality (AR) is mainly a developed tool for the visualization of 3D models and other relevant information overlaid in a real world scenario. Using this technology, we find previous studies about the relationship between student motivation, degree of satisfaction, and the user experience or student perception in the interaction with and teaching of applied collaborative works is extensive, with recent contributions that have helped to design new e-learning experiences or dislocated teaching using IT, and advanced visualization tools like AR [1], [2].

This technology is more extensively studied from a technological perspective (the Institute of Electrical and Electronics Engineers (IEEE) International Symposium on Mixed and Augmented Reality (ISMAR) is the global reference in these advances) or from the perspective of sociological and communication impacts (as addressed by the annual conference of the International Communication Association) instead of its educational capacity or ability to transform teaching and education.

From the first experiences using this technology [3], we find different works that proposes a prototype that helps users to interact with the world, and more recent



proposals for users in their everyday interactions with the world [4], which shows a device that provides real-time information to the user.

The 3D models are being used in the field of architecture for the visualization of projects for years [5]. The incorporation of AR to these types of projects [6] has increased expectations of the use of this tool reaching to situate the 3D model in the place where it will be build [7].

This capacity of AR technology, which shows a "completed" reality superimposed on reality, allows for the creation of an impossible image of what does not exist as a result of the analysis of existing building systems (e.g., structural, facilities, and envelope) and geo-location and photo composition. AR could facilitate rehabilitation and maintenance tasks, systems verification, and interactive updates in the same place and in real time, promoting more efficient management and control processes of building construction elements [8].

All of these improvements in space visualization and interpretation have clear relevance to the professional world and lead to a teaching process that allows for the rapid assimilation of concepts by the student [9].

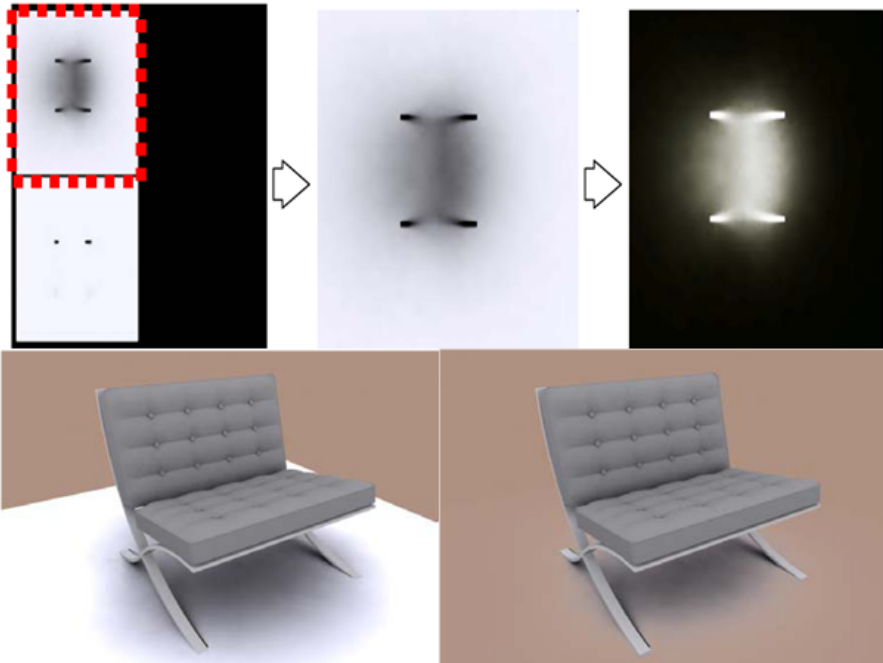
The main objective of the project is generate a simply dynamic lighting control based on a data-glove with accelerometers to interfere in advanced 3D visualization software used in the architecture framework to evaluate the behavior of shadows in 3D project models used in architecture education.

## **2 Using AR and Advanced Lighting in Educational Environment**

One of the challenges of visualization in these types of projects is lighting. Previous experiences have been developed with students in the visualization of objects that include lighting as a resource to achieve a better light immersion. The problem of virtual models illumination and how it can be integrated into the scene has been also widely discussed. In the first approaches to RA, the virtual object was simply overlapped in the real environment.

Major advances in technology focused on the correct calibration and registration of objects, studying the possible effects of occlusion and spatial coherence of objects, regardless of any other adaptation of the object in the scene. In other words, once the object was included in the scene, it was an artificial object, unable to adapt to the changes in environmental light. That kind of configurations lacked realism, and consistency of the scene was based only on geometrical aspects [10].

The sensation of realism in the scene is obtained primarily through visual interactivity. While it is true that as more senses involved, a greater sense of realism is achieved, a realistic immersion system should be able to create a complete visual simulation or as close as possible to it "Fig. 1".



**Fig. 1.** Texture maps to cast shadows in real space. On model basis, a lightmap is assigned as the main texture, and its inverse image is assigned as an opacity map to acquire transparency. So black pixels remained transparent, leaving visible only the cast shadow area.

Despite of all these improvements, the uses may lose the attention of the presentation during short necessary breaks needed for the presenter to interact with the computer for activating necessary commands (such keystrokes or mouse movements) for generating the desired changes to the model.

From an educational point of view, it is proved [11] that removing or minimizing the impact of these breaks can avoid the loss of attention from audience/students. Previous experiences [12], aims to improve comprehension of the project presented.

For this reason our department (Architecture La Salle, Ramon Llull University in collaboration with Graphical Expression Department of the Polytechnic University of Catalonia – UPC) has an active open line of research about the impact of technology on improving the understanding of the information presented.

As a part of that line of research we are working on a project that allows the possibility of interaction with the hardware used for showing the information in a real-time intuitive way.

One of the hypotheses of the project is to quantify the change in the dimension of the project presentations.

To allow this interaction, we have embedded electronic components such as accelerometers and sensors to devices like “data gloves”, commands, models or objects related to the project. This offers a wide range of possibilities in constant development up to date, making the presentation more spectacular.

These devices allow independence and self-reliance during presentation, so that the transmission of the information can be better focused on the audience as revealed by other studies [13].

In this case an application has been implemented for acting in the software AR-Media Plugin® of Inglobe Technologies® for Autodesk® 3DS MAX.

The main thing of the software developed consists in simulating the keystrokes and mouse movements which controls the AR-Media Plugin depending on the information received from the device controlled by the presenter.

Thus, with this movements acquired by the “data-glove” or model we can modify the parameters of the light source, its path or even show additional content of the project with no need to approach at the computer.

We are currently working on the first of the four phases in which the project is formed:

- **1<sup>st</sup> Phase: Project Definition:** Defining the problem; Proposed solution; Implementation; Study hypothesis.
- **2<sup>nd</sup> Phase: Study.** First tests; Data acquisition.
- **3<sup>rd</sup> Phase: Analysis.** Analysis of results; Hypothesis review; Proposed improvements.
- **4<sup>th</sup> Phase: Improvements.** Implementation; Test.

### 3 Materials and Methods

The lighting control is performed by an application in Visual Studio.net (VS.net) that is able to acquire data from external devices such as sensors, transducers and accelerometers, as in other references [14]. The developed system provides an effective solution for the data collection system in real practice [15].

This particular project has been implemented in two versions, the wired version (A) and the wireless version (B) with same functionality but different features. In both cases we use the triple-axis ADXL335 accelerometer.

#### 3.1 Wired Version (A)

The accelerometer is embedded inside a polystyrene sphere “fig. 2”, simulating the Sun as light source. With NI-DAQ6009 from National Instruments® supplies 2.5V and capture input analogic signals in AI0, AI1, AI2.

This system proposed allows scanning full scale and captures the sensibility of the movement by the user changing the position of the light source. Once we have the signal digitalized and quantized a basal value is fixed corresponding to a neutral position of the accelerometer.

From this point, the movement of the data-glove modifies the output signal corresponding to x, y, z axis.

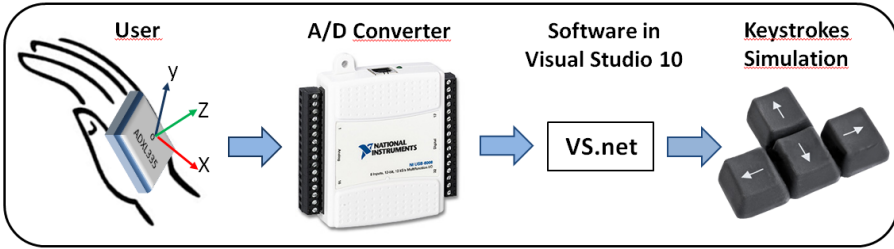


Fig. 2. Application architecture of version A

The developed software associates each value of the NI-DAQ6009 inputs with light controls of the plugin for modifying the position of the light source.

For moving the light source to the right, the software simulates the keyboard key “→”, move left “←”, move forward “↑” and move backwards “↓”. To move upwards in z axis the user needs to push the following key combination: “Ctrl+↑” and “Ctrl+↓”. The “Fig. 3” shows the diagram of the dataflow generation and acquisition.

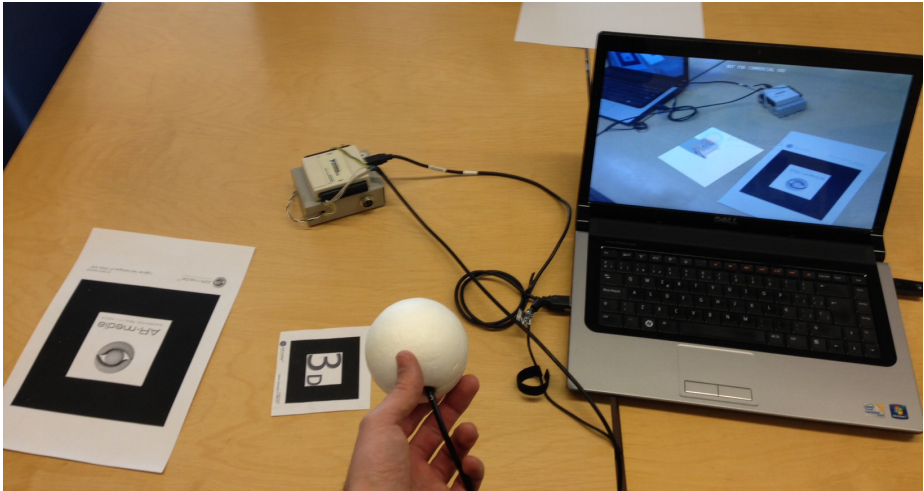
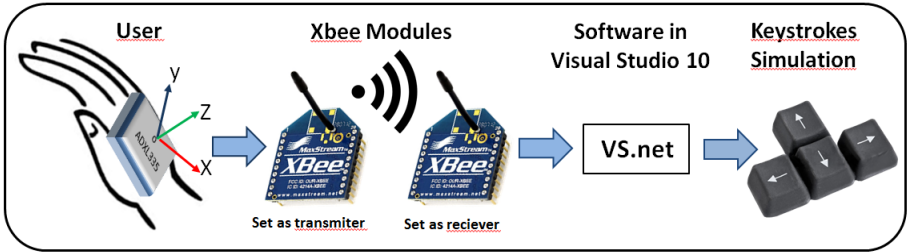


Fig. 3. Image of the implementation with the wired polystyrene sphere and the NI-DAQ6009

### 3.2 Wireless Version (B)

In this case, two XBee® RF Modules are used [16]. These Modules, the 3-axis accelerometer and the breakout board FT232 RL USB to Serial are weld and configured properly “Fig. 4”.



**Fig. 4.** Application architecture of version B

When connecting any analogical output of the accelerometer to any digital input of the XBee® set as a transmitter, it is important to know the behavior of the data transfer system. If the value of the signal from the accelerometer exceeds 2,5V, the behavior on the digital input it is like writing a “1” and when the value is lower than 2,5V, means writing a “0”.

This data package is received by the other XBee® module set as a receiver and transmitted to the PC via the breakout board FT232 RL USB to Serial port. The software reads this data package as data-glove movements and according to these values the keyboard keys are simulated “Fig. 5” in the plug-in AR-Media® in 3DS MAX® environment.

```

If X_Axis > 2.6 Then
    SendKeys.Send("{DOWN}")
End If
If X_Axis < 2.1 Then
    SendKeys.Send("{UP}")
End If
If Y_Axis > 2.6 Then
    SendKeys.Send("{RIGHT}")
End If
If Y_Axis < 2.2 Then
    SendKeys.Send("{LEFT}")
End If

```

**Fig. 5.** Commands in VS.net for simulating keyboard keys for  $x$  and  $y$  axis of the accelerometer. The same for  $z$  axis

## 4 Conclusions

The results of the design of the project make us to fix the target in the following concept: how is it possible to increase the ease of use in dynamic lighting control based on a data-glove with accelerometers or other human interface and the ability to interfere in advanced 3D visualization software with simple lines of code.

The partial control of the 3D software with two versions with the same functionality notes that the design of the project developed improves handling and speeds up

user interaction. Furthermore, the real-time gestures read by the device give more realism and makes the user immerse itself into the project.

This project has been developed in the department of Architecture La Salle, Campus Barcelona, Universitat Ramon Llull. The aim of the research is to find new ways to explore interaction of users with projects of architecture. This study case goes deep into the lighting processes of the architecture illumination. The interaction in real time could improve the strategies to project the shapes of architectural designs of the buildings in order to get profit of the solar radiation and being more energy efficient. This approach to solar studies will generate results which will be one way to reach a sustainable architecture thanks to Augmented Reality.

The next phase (modeled by the CAD/BIM/AR group of the same faculty), will be performed during the 2012-2013 academic year with students in their fourth year of an Architecture and Building Engineering degree. The experimental framework is in progress in the course “Sustainability and Energy Efficiency,” a nine-ECTS-credit course that is taught in the second semester.

In summary, this project presents a smart way to interact on very powerful software packages allowing emulate its commands from an external device equipped with sensors, accelerometers or other components integrated in data-gloves or data-suits. Simulating control commands with few lines of code enhance the presentation to a higher level. The solution tested in this project with 3DS MAX® and the AR-Media Plugin®, can be extrapolated to almost all 3D modeling and AR programs.

Next step in Phase 2 of the project is the evaluation with users to obtain results in order to study the first design and possible changes to improve the system.

## References

1. Sun, J., Hsu, Y.: Effect of interactivity on learner perceptions in Web-based instruction. *Computers in Human Behavior* 29(1), 171–184 (2013), doi: 10.1016/j.chb.08.002
2. Giesbers, B., Rienties, B., Tempelaar, D., Gijssels, W.: Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing. *Computers in Human Behavior* 29(1), 285–292 (2013), doi: 10.1016/j.chb.2012.09.005
3. Feiner, S., Macintyre, B., Seligmann, S.: Communications of the ACM - Special Issue on Computer Augmented Environments: Back to the Real World 36(7), 53–62 (1993)
4. Google Project Glass ® (2013), <http://www.google.com/glass>
5. Brooks Jr., F.P.: Walkthrough—a dynamic graphics system for simulating virtual buildings. In: *Proceedings of the 1986 Workshop on Interactive 3D Graphics*, Chapel Hill, North Carolina, USA, pp. 9–21 (January 1987)
6. Kim, H.: A bird’s eye view system using augmented reality. In: *32nd Annual Simulation Symposium in Proceedings*, pp. 126–131 (1999)
7. Sánchez, J., Borro, D.: Automatic augmented video creating for markerless environments. Poster Proceedings of the 2nd International Conference on Computer Vision Theory and Applications, VISAP 2007, pp. 519–522 (2007)
8. Sánchez, A., Redondo, E., Fonseca, D., Navarro, I.: Construction processes using mobile augmented reality. A study case in Building Engineering degree. In: *World Conference on Information Systems and Technologies (WorldCIST 2013)*, Algarve, Portugal, March 27-30 (2013)

9. Vechhia, L.D., Da Silva, A., Pereira, A.: Teaching/learning Architectural Design based on a Virtual Learning Environment. *International Journal of Architectural Computing* 7(2), 255–266 (2009), doi:10.1260/147807709788921976.
10. Redondo, E., Fonseca, D., Sánchez, A., Navarro, I.: Augmented Reality in architecture degree. New approaches in scene illumination and user evaluation. *International Journal of Information Technology and Application in Education (JITAE)* 1(1), 19–27 (2012)
11. Kalkofen, D., Mendez, E., Schmalstieg, D.: Comprehensible visualization for augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 15, 193–204 (2009)
12. Fonseca, D., Marti, N., Navarro, I., Redondo, E., Sanchez, A.: Using augmented reality and education platform in architectural visualization: Evaluation of usability and student's level of satisfaction. In: *International Symposium on Computers in Education (SIIE)* (2012)
13. Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A., Oberman, M., Teitel, M.: Reality built for two: a virtual reality tool. In: *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, Snowbird, Utah, USA, pp. 35–36 (February 1990)
14. Fortin, P., Hebert, P.: Handling occlusions in real-time augmented reality: dealing with movable real and virtual objects. In: *The 3rd Canadian Conference on Computer and Robot Vision, CRV 2006*, p. 54 (2006)
15. Yan, H., Kou, Z.: The design and application of data acquisition system based on NI-DAQ. College of mechanical engineering, Taiyuan, China (2010)
16. ZigBee Technology, <http://www.digi.com> (retrieved on March 2012)

# Ultra Low Cost Eye Gaze Tracking for Virtual Environments

Matthew Swarts<sup>1</sup> and Jin Noh<sup>2</sup>

<sup>1</sup>Georgia Institute of Technology, College of Architecture, Atlanta, GA, USA  
matthew.swarts@coa.gatech.edu

<sup>2</sup>Gwinnet School of Mathematics, Science, and Technology, Lawrenceville, GA, USA  
noh.jin2014@gmail.com

**Abstract.** In this paper we present an ultra-low cost eye gaze tracker specifically aimed at studying visual attention in 3D virtual environments. We capture camera view and user eye gaze for each frame and project vectors back into the environment to visualize where and what subjects view over time. Additionally we show one measure of calculating the accuracy in 3D space by creating vectors from the stored data and projecting them onto a fixed sphere. The ratio of hits to non-hits provides a measure of 3D sensitivity of the setup.

**Keywords:** Low Cost, Eye Tracking, Virtual Environments.

## 1 Introduction

In this paper we present a method for constructing an ultra-low cost eye gaze tracking system aimed at use in visual attention studies within 3D virtual environments. The motivation behind the development of low cost eye gaze tracking systems for virtual reality lies in the use of spatial analysis, human behavior research, and neuroscience to uncover how the structure and material of a space affects the understanding and cognition of space. Virtual reality has been used in the field of space syntax [1] to study the paths people take in a new environment. Eye tracking could be used to look deeper into the motivating factors of features of the space that entice people to take specific directional queues. Virtual reality has also been used in neuroscience [2] to replicate maze and visual puzzle experiments traditionally performed on mice. Eye tracking has been used in these instances as well. In museums, artifacts have been shown to be grouped both spatially as well as visually to create a rich set of connections [3]. These types of visual pairings and groupings could also be better understood through the use of eye gaze tracking. When the building no longer exists, the building was never constructed, the layout of the museum or building has changed through renovation, or for testing hypothetical spatial structures, eye tracking can be combined with virtual reality to gain understanding beyond what is available in the built environment.

Tracking eye movements allows us to see a fairly involuntary human response to an environment. Eye movements provide a more quantitative method of studying



human behavior and perception [4]. The eye moves in patterns of fixations and rapid saccades based on what is being viewed, when, and where. Capturing these movements allows us to see underlying structures in complex objects and patterns in the surrounding environment.

Eye tracking is used for a myriad of purposes including advertising and marketing, training, assistive technology, and psychological studies. Inspection and training both in the physical environment [5] and in 3D virtual environments [6, 7] can utilize eye tracking. Within the realm of virtual environments, eye tracking is used for object manipulation [8, 9] and user movement. It can also be used to show predicted eye movements based on visual attention cues [10]. Experiential effects, such as depth of field, can be improved using eye tracking [11]. A user's anticipation of a turn in active navigation of a virtual environment can be determined by observing eye movements [12]. Salient maps of features in an environment can be used to improve the accuracy of eye trackers [13-15] by applying attention theory [16]. Several devices also employ two eye trackers for binocular tracking to determine precise 3D location [6, 7, 17] and user movement [18].

Eye trackers are used for people with motor impairments or other disabilities, allowing them to interact with the physical and virtual environment [17, 19]. Along these lines, others have developed methods of reducing the size and weight of portable eye trackers [17, 20] and for making low-cost eye trackers [17, 21, 22] to increase the general accessibility of the technology.

## 2 System

Our system is composed of eye tracking hardware, eye tracking software, and a 3D virtual environment model with network messaging and analytics for processing and post-processing of the input data streams.

### 2.1 Eye Tracking Hardware

The eye tracking hardware is made from a camera, a lens, a filter, a clamp, a helmet mount, and infrared LEDs. The selection of each element was a balance among cost, availability, weight, and expected accuracy. The overall design was a helmet or head mounted eye tracker in which the image of the eye could be maximized for better accuracy. Other designs, such as a remote monitor mounted tracker, were considered, but the distance is an issue for maintaining spatial accuracy with the limited hardware.

Construction of an eye tracker generally requires a camera or two. For virtual reality that is presented on a single monitor, so only one camera is necessary for most setups. While binocular eye trackers exist for some virtual reality setups, they are not very useful for a single screen without true 3D capability. Tracking both eyes allows the capture of the user's focal plane. However, in using a single display screen, the focal plane can be assumed to be the screen itself. We selected the Sony PlayStation Eye camera, as it is possible to get speeds up to 187 frames per second (fps) at a resolution of 320x240 or 60fps at 640x480. The higher frame rate allows us to test higher temporal resolutions in capturing human eye movements than traditional web cameras

running at 30 frames per second. The high frame rate capability of the PlayStation eye was developed by Sony by the maximizing the use of the USB 2.0 bandwidth limits. This limitation of data bandwidth is what distinguishes most low cost commercial off the shelf product (COTS) USB 2.0 color web-cameras from higher-end more expensive (>\$500 USD) industrial application [23] single band cameras which use IEEE 1394 FireWire, Camera Link, or the new USB 3.0 specification with higher data transfer rates.

The camera was disassembled, and the plastic cover removed. The plastic lens holder was removed, and replaced by a lens mount with threading for m12 lenses. An infrared (IR) band-pass filter was inserted into the new camera lens mount to only allow IR light onto the imaging array. An 8mm optical lens was screwed into the new lens mount. This lens provided a larger, zoomed in view of the user's eye when mounted to the helmet, allowing more space for pupil view analysis, while keeping some distance from the user's eye.

Infrared light is typically used in eye tracking, because it is not in the visible spectrum and does not interfere with the user's vision. Additionally the human iris reflects infrared light, making the iris appear lighter regardless of the visible eye color.



**Fig. 1.** Camera assembly and helmet mount

This provides a way to easily discern the pupil from the iris. The pupil center can be more easily determined, which is a feature used for comparison in most eye tracking methods. An infrared light source was constructed from IR LEDs to illuminate the eye for use in indoor environments.

There are several options when it comes to mounting the camera assembly for monitoring computer display interaction. The typical setups include mounting to the bottom of the monitor as a remote camera, and mounting to the user's head using glasses or a helmet. We chose a head mounted approach using a helmet in order to increase the spatial resolution of the eye movements and to minimize calibration issues associated with large zoom lenses. Metal alligator clips on rods were taken from an electronics magnifier and used to hold the camera assembly onto the helmet while allowing for the slight adjustments needed for different users.

## 2.2 Eye Tracking Software

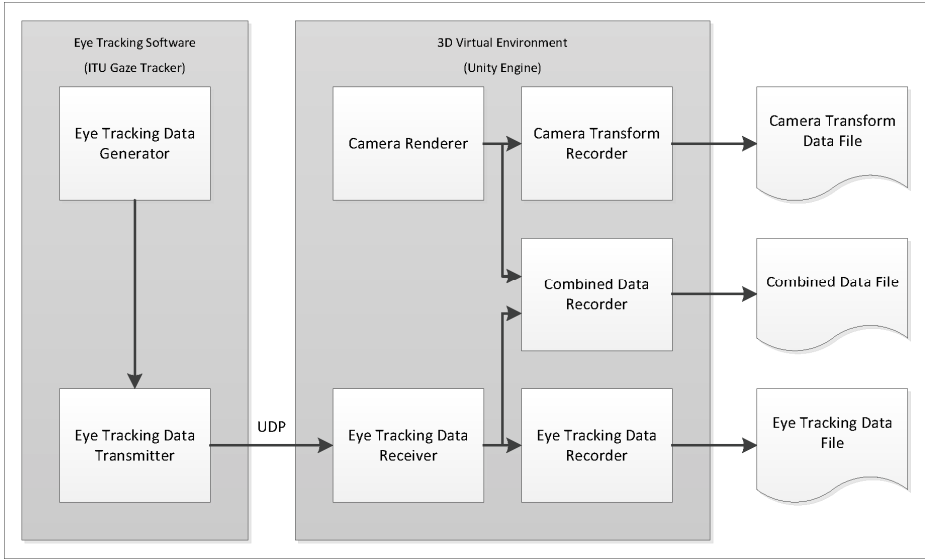
The core of the system relies on the work of the ITU GazeGroup [24]. The ITU Gaze Tracker is an open source eye tracker developed at IT University of Copenhagen [22] aimed at providing low cost alternatives to commercial systems and making it more accessible. The software is extremely flexible in terms of input hardware, hardware setup, and feature tracking. Their system also allows for control over the sensitivity of each feature. Additionally, there are several levels of calibration available. The element that is most important for our system is the ability to stream out the view location in screen coordinates over the network via Universal Datagram Protocol (UDP).

In addition to the ITU Gaze Tracker software, we also used drivers developed by Code Laboratories [25] specifically for the Sony PlayStation 3 Eye Camera. These drivers provide access to the higher frame rates available through the camera.

## 2.3 Virtual Environment

The Unity Game Engine [26] is a popular 3D video game engine used for developing and publishing video games on many platforms. It is free for non-commercial use, and extremely flexible with three powerful scripting languages, and the ability to bind to external libraries. We developed a set of scripts to capture the view points as well as the position, location, and field of view of the user along with the current system timestamp. The data is saved to a file, which can be loaded, analyzed, and visualized.

The user's view is captured by the ITU Gaze Tracker, which sends the coordinates to Unity via UDP. During each frame of the virtual environment, the position, rotation, and field of view of the user's camera is captured. If there has been a new view coordinate received since the last frame, then the 2D screen coordinates from the ITU Gaze Tracker are concatenated with the 3D camera data, and the entire data set is stored to file. This ensures that there is always 3D camera data associated with a 2D view coordinate. For this association we assume that the virtual environment operates at a higher frequency than the eye tracker. In our tests, this was the case, as the environments were highly optimized, and the limiting factor was the frequency of the eye tracking camera. However, we do record both the camera data and the eye tracking data separately as well for cases where a custom post-synchronization step is necessary.



**Fig. 2.** Overall software architecture. Eye tracking software passes eye tracking data through UDP. The 3D Virtual Environment receives the eye tracking data and records it to a file. It also captures the current camera matrix and saves the 3D position and view data to a file.

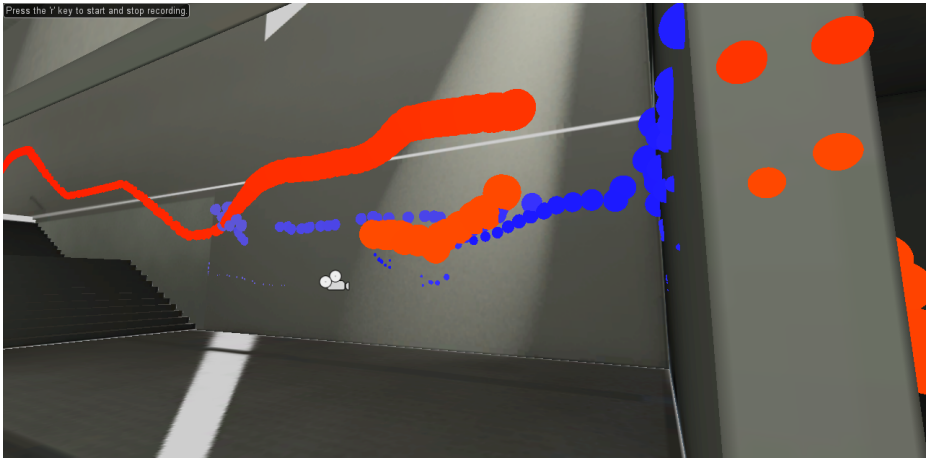
## 2.4 Visualization

One typical method for analyzing eye tracking data in two dimensions is through the use of a heat map. A heat map of eye movements or mouse movements is an aggregation of time spent in each location of the 2D space. This is accomplished by applying a circle with a radial gradient of transparency to each recorded location with an intensity of the duration spent in that location. This 2D visualization method is not particularly well suited to 3D space without much more data points to aggregate.

The visualization in Figure 3 is a 3D virtual environment constructed using Autodesk 3D Studio Max and Adobe Illustrator. Our initial motivations prompted the use of virtual museums for initial testing. As an example we selected Tadao Ando's Pulitzer Foundation of the Arts, for its ability to produce alternative visual and spatial interpretations using space, light, and color [27].

To visualize the eye tracking data in 3D space we use the camera projection matrix, which includes the position, rotation, and field of view. We then take the 2D view coordinate data and using the recorded dimensions and aspect of the display, we calculate a projection vector. Using a screen-to-world operation we project the vector into the 3D environment from the camera location until it hits a surface. At that hit location in 3D space, we create a colored sphere. The center of the sphere corresponds to where the user was looking at that point in time. The radius of the sphere corresponds to the distance between the camera and the hit location. Smaller spheres represent smaller viewing distances, while larger spheres represent larger viewing

distances. Lastly, the color ranges between a spectrum of two hue values in the Hue-Saturation-Value (HSV) color space to represent the time within the trial.

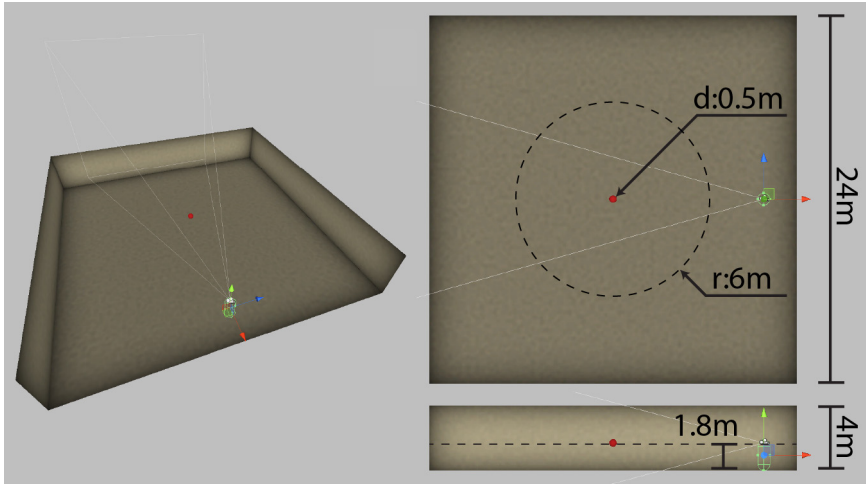


**Fig. 3.** Screen shot of a visualization of the eye tracked data over one user session in a 3D virtual environment

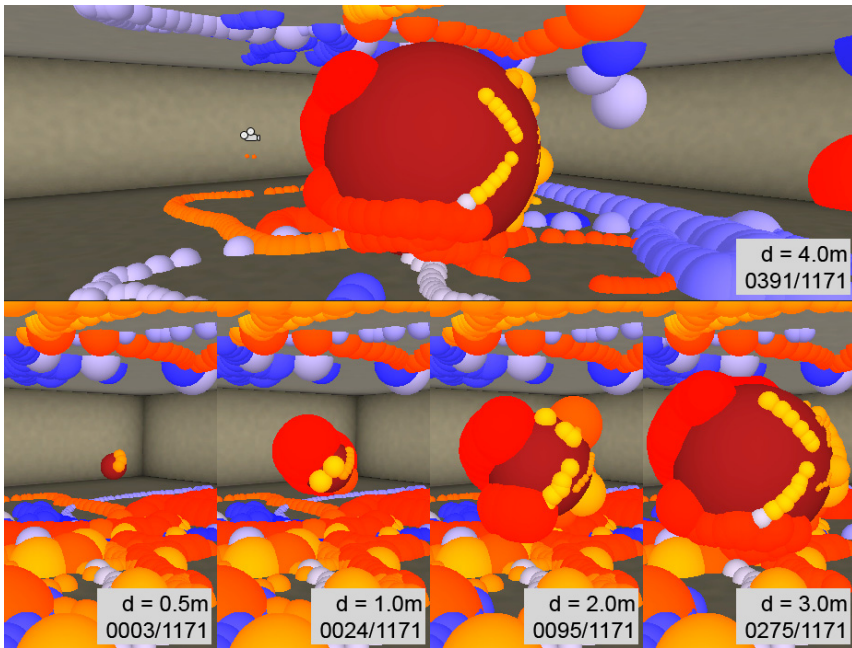
### 3 3D Accuracy Testing

As a preliminary test for accuracy, we designed a test environment in which the subject is asked to focus on a specific object. The environment, see Figure 4, consists of a rectangular room 24m wide by 24m in depth and 4m in height. A red sphere of 0.5m diameter is placed at the center of the room. The camera is set at an eye height of 1.8m. As the subject moves around the space, keeping the red sphere in view, the user is asked to keep their eyes on the sphere.

After a user has moved through the space, viewing the center sphere, we are able to project rays from each recorded view to a location in the space. By varying the size of the center sphere, we are able to see and count what ratio of hits contacted with the sphere, and which did not. Since the user continuously views the center sphere, the trajectory of the ray should tend to intersect the sphere, but with variance due to the accuracy of the eye tracking, and of the ability of the subject to remain in position. Figure 5 shows a screen capture of projections from 1 subject trial, using 5 different size diameter center spheres. The ideal case is that all of the hit spheres would be attached to the center sphere when its diameter is equal to 0.5m, the same as in the experiment. As we increase the sphere size, we get a sense of the error falloff in three dimensions.



**Fig. 4.** The layout and dimensions of the test environments with a sphere located at the center of a large rectangular room



**Fig. 5.** A sequence of screen captures of projected spheres of increasing size for one subject trial. Larger spheres identify a larger error capture zone. 391 out of 1171 samples were captures by a sphere of diameter 4.0m.

## 4 Conclusions and Future Work

We have demonstrated that an eye tracker for virtual reality can be developed at low cost, which can be used for some areas of research within 3D virtual environments. Additionally we provided a new measure for 3D sensitivity to indicate the accuracy of the system for a particular eye tracker setup.

Future work could incorporate an open-source model of 3D printed glasses specifically for the PS3 Eye Camera. Additionally the integration of salient maps to lock onto the most likely candidates for visual attention, may be useful for better accuracy in low cost systems. Lastly more measures of 3D movement, including angular velocities and angular accelerations, could be incorporated into a more robust regression model to determine which aspects of interaction with the 3D virtual environment are likely to cause the most errors in hit detection for low cost systems.

## References

1. Dalton, R.C.: The secret is to follow your nose: route path selection and angularity. *Environment & Behavior* 35, 107–131 (2003)
2. Moffat, S.D., Resnick, S.M.: Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behavioral Neuroscience* 116, 851–859 (2002)
3. Zamani, P.: Views across boundaries and groupings across categories (electronic resource): the morphology of display in the galleries of the High Museum of Art 1983–2003. Georgia Institute of Technology, Atlanta (2008)
4. Yarbus, A.L.: Eye movement and vision. In: Haigh, B. (trans.) Plenum Press, New York (1967)
5. Megaw, E.D., Richardson, J.: Eye movements and industrial inspection. *Applied Ergonomics* 10, 145–154 (1979)
6. Gramopadhye, A.K., Melloy, B.J., Nair, S.N., Vora, J., Orhan, C., Duchowski, A.T., Shivashankaraiah, V., Rawls, T., Kanki, B.: The use of binocular eye tracking in virtual reality for aircraft inspection training. *Int. J. Ind. Eng.-Theory Appl. Pract.* 9, 123–132 (2002)
7. Duchowski, A.: *Eye tracking methodology: Theory and practice*. Springer (2007)
8. Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: *Proceedings of the 1997 Symposium: Interactive 3D Graphics*, vol. 35 (1997)
9. Tanriverdi, V., Jacob, R.: Interacting with Eye Movements in Virtual Environments, pp. 265–272. Association for Computing Machinery (2000)
10. Peters, R.J., Itti, L.: Computational Mechanisms for Gaze Direction in Interactive Visual Environments, pp. 27–32. ACM (2006)
11. Hillaire, S., Lecuyer, A., Cozot, R., Casiez, G.: Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In: *Virtual Reality Conference, VR 2008*, pp. 47–50. IEEE (2008)
12. Hillaire, S., Lecuyer, A., Breton, G., Corte, T.R.: Gaze behavior and visual attention model when turning in virtual environments. In: *Proceedings of the 16th ACM Symposium: Virtual Reality Software & Technology*, vol. 43 (2009)
13. Hillaire, S., Breton, G., Ouarti, N., Cozot, R., Lecuyer, A.: Using a Visual Attention Model to Improve Gaze Tracking Systems in Interactive 3D Applications. *Comput. Graph. Forum* 29, 1830–1841 (2010)

14. Hillaire, S., Lecuyer, A., Regia-Corte, T., Cozot, R., Royan, J., Breton, G.: Design and Application of Real-Time Visual Attention Model for the Exploration of 3D Virtual Environments. *IEEE Trans. Vis. Comput. Graph* 18, 356–368 (2012)
15. Lee, S., Kim, G.J., Choi, S.: Real-Time Tracking of Visually Attended Objects in Virtual Environments and Its Application to LOD. *IEEE Trans. Vis. Comput. Graph* 15, 6–19 (2009)
16. Treisman, A., Gelade, G.: A feature-integration theory of attention. In: Wolfe, J., Robertson, L. (eds.) *From perception to consciousness: Searching with Anne Treisman*, pp. 77–96. Oxford University Press, New York (2012)
17. Abbott, W.W., Faisal, A.A.: Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain-machine interfaces. *J. Neural Eng.* 9 (2012)
18. Munn, S.M., Pelz, J.B.: 3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker. *Eye Tracking Research & Application* 181 (2008)
19. Adjouadi, M., Sesin, A., Ayala, M., Cabrerizo, M.: Remote eye gaze tracking system as a computer interface for persons with severe motor disability. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) *ICCHP 2004*. LNCS, vol. 3118, pp. 761–769. Springer, Heidelberg (2004)
20. Babcock, J.S., Pelz, J.B.: *Building a Lightweight Eyetracking Headgear*, pp. 109–114. Association of Computing Machinery, New York (2004)
21. Kassner, M., Patera, W.: PUPIL: constructing the space of visual attention. Dept. of Architecture, vol. Masters, pp. 181. Massachusetts Institute of Technology (2012)
22. San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D.W., Hansen, J.P.: Evaluation of a low-cost open-source gaze tracker. *Eye Tracking Research & Application* 77 (2010)
23. Point Grey Research, <http://ww2.ptgrey.com/>
24. ITU GazeGroup, <http://www.gazegroup.org/>
25. Code Laboratories > CL Studio Live, <http://codelaboratories.com/>
26. Unity - Game Engine, <http://unity3d.com/>
27. Bafna, S., Losonczi, A., Peponis, J.: Perceptual Tuning of a Simple Box. In: *Eighth International Space Syntax Symposium* (2012)



# Real-Time Stereo Rendering Technique for Virtual Reality System Based on the Interactions with Human View and Hand Gestures

Viet Tran Hoang, Anh Nguyen Hoang, and Dongho Kim

Soongsil University  
511 Sangdo-dong Dongjak-gu, South Korea  
{vietcusc, anhnguyen}@magiclab.kr,  
cg@su.ac.kr

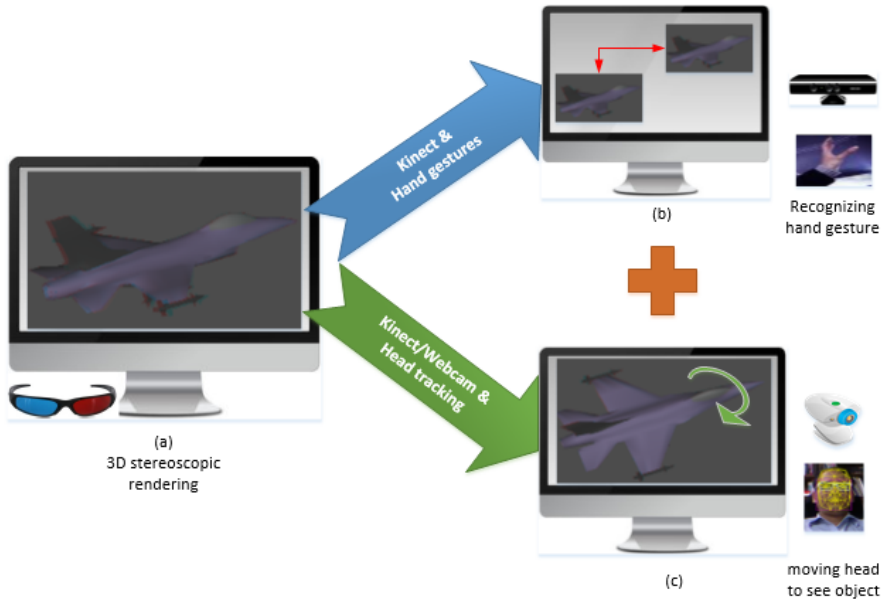
**Abstract.** This paper proposes the methods of generating virtual reality system with stereo vision, simple and widely used 3D stereoscopic displays. However, we are motivated by not only 3D stereo display but also realistic rendered scenes popped out of screen which can be thought of as an interactive system addressing the human-to-virtual-objects manipulation. The user of the system can observe the objects in the scene in 3D stereoscopy and can manipulate directly by using hand gestures. We present the technique to render the 3D scene out of the screen and use KINECT device to keep track of user's hand movement to render the objects according to the user's view.

**Keywords:** virtual reality, stereoscopy, real-time rendering, head tracking.

## 1 Introduction

Our system presents a stereo rendering technique by which virtual models are superimposed on the computer screen and appear as realistic models in the human interactions. We present an algorithm to compute and assign the parallax value to the pair of left and right stereo images of an object in the screen space to "bring" the objects out of the screen in a whole real-time process respecting the consistence of disparity map of the original 3D rendered scene. We limit the position of the user at the distance of 40-60cm to the screen in order that the user can conveniently reach and manipulate the objects virtually.

In a stereo rendering system, it is highly required to maintain the position and orientation of the user's view toward the screen display as a straight gaze; therefore the human perception of depth and immersion can be kept stable and most accurate. Any changes in the viewer's position and direction can always potentially result in the distortion of the scene or the objects in screen space and lead the human perception into some negative symptoms of the eyes such as eye strain or fatigue. In our system, we track the human head pose, and from the information collected, we introduce a new projection matrix calculation to adjust the projection parameter for the rendering of a new scene so that the viewer cannot feel (or at least cannot perceive easily) the distortion of the scene while the 3D stereo vision can retain the fidelity.



**Fig. 1.** Using KINECT and webcam to help the viewer to see object more realistic and interacting with them in 3d stereoscopic environment

In many cases, exposure to stereoscopic immersion techniques can be lengthy so that the user can face eye strain. The more the viewer's eyes are not oriented straight ahead to the computer screen, the more distortion of the objects will occur dramatically. We limit the angle of the user's orientation to the computer screen by 45 degree from both left and right sides from the straight direction in order to keep the rendered scene staying realistic for human perception.

In addition to enhance the interaction, we develop some basic manipulations that the viewer can perform on the virtual scene and objects. We use the KINECT device to capture the hand gestures of the viewer. We define and implement some new hand gestures based on the fundamental implementation of KINECT SDK. The viewer can perform gestures while recognizing the change of rendered scene. We define some signs of these changes in the scene objects such as they appear marked with colors when being touched; they are moving or rotating corresponding to the movement of hands.

## 2 Related Work

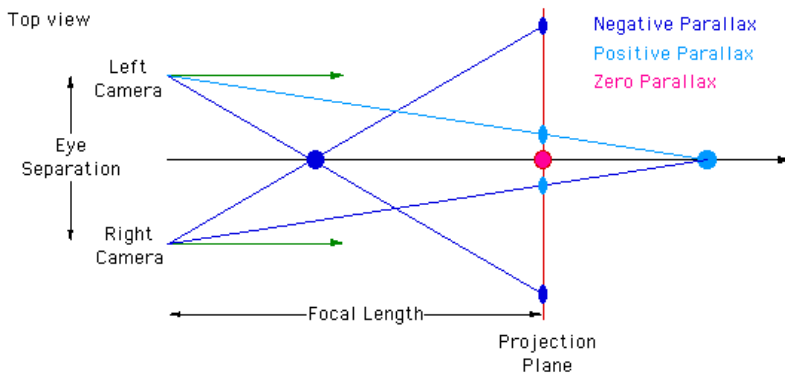
Using 3D stereoscopy in movie is a new trend in the world. There are a lot of famous movies that are produced by using 3D technology. And, some producers have spent money and time to convert their movies from 2D to 3D products. There are many researches in stereoscopic 3D to apply in many fields.

Paul Bourke wrote “Calculating stereo pairs” [10] in July, 1999 for discussing on the generation of stereo pairs used to create the perception of depth that are very useful in many fields such as scientific visualization, entertainment, games, etc. In his researches, the author was using stereo pairs with one of the major stereo 3D display technologies to create virtual three dimensional images. He already calculated eye separation distance and focal length to define category of parallax: positive parallax, negative parallax and zero parallax.

Paul Bourke created anaglyphs using OpenGL [10] in 2000 and updated these anaglyphs in 2002 by using GLUT library to do filtering automatically for left and right eye images. By using OpenGL, another person - Animesh Mishra was also rendering 3D anaglyphs [11] more precisely and effectively than previous one. He measured the amount of parallax for a vertex beyond convergence distance, calculated distance between intersections of left eye, right eye with screen for each case of parallax.

In 2008, François de Sorbier, Vincent Nozick and Venceslas Biri presented GPU-based method to create a pair of 3D stereoscopic images [7]. This is a new method using the advantages of GPU to render 3D stereo pairs including geometry shaders.

Besides, KINECT’S SDK tools support some tracking methods for using this device such as hands gesture capturing. Jens Garstka and Gabriele Peters demonstrated a view-dependent 3D projection using depth image based Head tracking method. They discussed about how to use depth image algorithm when they tracked a head. In this method, they used the depth images to find the local minima of distance and surrounding gradients to identify a blob with the size of a head, then transformed the drawn data and processed these data for tracking of a head. With view-dependent 3D projection, it provided the viewer a realistic impression of projected scene regarding to his/her position in relation to the projection plane.



**Fig. 2.** Definitions in anaglyphs technology

### 3 Theory and Concept

#### 3.1 3D Stereoscopy

In this paper, we focus on rendering 3D stereo using anaglyphs technology. How to render stereoscopy? In order to render stereo pairs, we need to create two images, one for each eye. We must understand some definitions about parallax, eye separation, aperture, etc.

For parallax, the distance between the left and right eye projection is called the horizontal parallax. If the object is in the opposite side from the eyes over the projection plane, it is called positive parallax. If the object is located in front of the project plane and same side with eyes, it is called negative parallax. Final definition is zero parallax where the object is located right on the projection plane.

To generate the stereoscopic images or objects on the screen, we need two images: one for left eye and one for right eye. There are two general approaches to make these images: Toe-in and Off-axis. Toe-in makes the viewer feel sick or gives some sorts of headache while Off-axis does not cause any problems. Off-axis approach is the better one and it also uses two asymmetric frustums. To get two pictures for left and right eyes, we need three steps: transforming camera (translation), calculating frustums and rendering of scene.

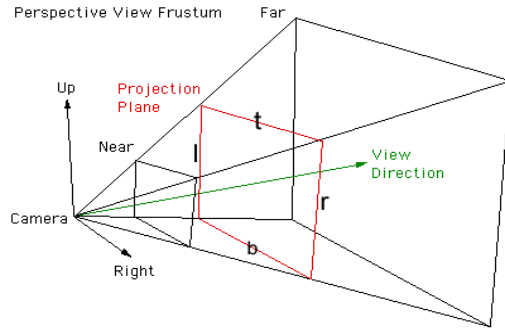


Fig. 3. Perspective view frustum

$$wd2 = \text{near} \cdot \tan\left(\frac{\pi}{180} \cdot \frac{FOVy}{2}\right) = t; \tag{1}$$

$$b = - wd2 \tag{2}$$

\* Frustums for left:

$$l = b \cdot \frac{\text{width}}{\text{height}}; \tag{3}$$

$$r = b \cdot \frac{\text{width}}{\text{height}} + 0,5 \cdot \text{eye\_sep} \cdot \frac{\text{near}}{\text{focal}} \tag{4}$$

\* Frustums for right:

$$r = b \cdot \frac{\text{width}}{\text{height}}; \tag{5}$$

$$l = b \cdot \frac{\text{width}}{\text{height}} + 0.5 \cdot \text{eye\_sep} \cdot \frac{\text{near}}{\text{focal}} \quad (6)$$

As mentioned above, GPU-based geometry shaders can be used to render 3D stereoscopy by processing vertices and pixels. The main purpose of geometry shaders is to clone input primitives without requiring any process on the vertex attributes while traditional method renders it twice. By using the power of graphic cards, the performance of 3D stereo rendering in GPU-based method is approximately faster twice than the traditional methods. This method is very useful for rendering 3D stereo anaglyphs.

### 3.2 KINECT Tracking

There are so many researches in head pose tracking and gesture tracking. One of tracking technique is to calculate depth-images. In 2010, Microsoft launched the game controller KINECT with Xbox 360. The basic principle of KINECT'S depth calculation is based on stereo matching. It requires two images: one is captured by the infrared camera, and the other is the projected hard wired pattern [6]. These images are not equivalent because some distances between camera and projector. Therefore, we can calculate object positions in space by the view dependent 3D projection.

## 4 Implementation

With the expectation to reduce eye strain headache or sickness for the viewer, we implemented an application for 3D anaglyphs based on off-axis approach. Besides, to improve the performance of 3d stereoscopy, we also apply GPU-based shaders in the implementation.

### Step 01: Anaglyphs using GPU

To build 3D anaglyphs, we are following the concepts mentioned in the previous part and information in "Build your own 3D display" course [3]. In this part, we define the samplers corresponding to the left and right images, then use the geometry shaders to calculate in fragment and vertex shaders of GPU and assign the output fragment color to the anaglyphs rendering in the application. Belonging to the output of left and right matrices, we have different anaglyphs mode such as full color mode, half color mode and optimized color mode. The matrix of full-color anaglyphs is shown as below:

#### Color Anaglyphs

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

- Partial color reproduction
- Retinal rivalry

The result of this part is an application with an object in 3D stereo belonging to anaglyphs approach. The object can be scaled, changed the position or mono mode from 3d stereo modes. (Fig.5.a)

### Step 02: Tracking with KINECT

In our system, we use the 3GearSystems [9] in order to track hand gestures. This system enables the KINECT to reconstruct a finger-precise representation of hands operation. Therefore, this system gives the best results for hand-tracking process and allows us to integrate with the stereoscopic application.

3GearSystem can use either OpenNI [8] or KINECT SDK for Windows as the depth sensing camera SDK. In our system, we choose OpenNI SDK with two KINECT devices. 3GearSystem has many advantages in comparison to other systems used for hand gestures tracking. There are lots of KINECT software and algorithms working best when capturing large objects or full-body of the user. It is required that the user must stay away from KINECT sensor several meters. 3GearSystem uses two KINECTS to capture both hands and to enhance the precision of tracking process. The KINECT devices are mounted over a meter above the working place. This is to help the users work as in normal condition and they can perform hand gestures in front of their screens conveniently. This 3GearSystem satisfies our requirements.

3GearSystem uses a hand-tracking database to store the user's hand data. In order to integrate 3GearSystem into our application, we need to first calibrate the system and train 3GearSystem about our hands data. This training process consists of the following steps: calibration, training hands shape (Fig.4), training six hands poses, and creating the user data.

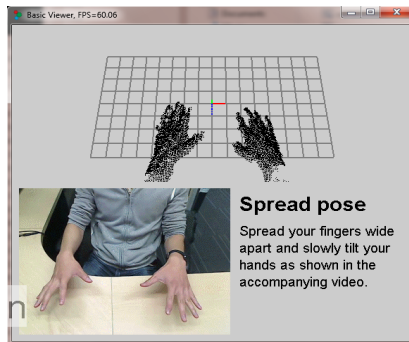


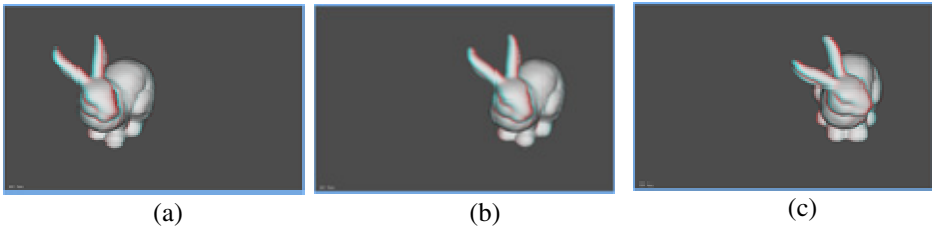
Fig. 4. Hand shape training

### Step 03: Integration of the stereoscopic rendering and tracking

This paper focuses to make the viewer comfortable when they are watching the 3d stereoscopy. Furthermore, the viewer will be more interested if they interact with objects in the system or when they are moving some things by hand gestures.

We also integrated hand gestures and head tracking. With hand gestures, viewer can move an object in the system from this place to another place or from back to front of the project plane. On the other hand, head tracking will track the viewer's head when it moves left-right or up-down and displays the hidden part of an object.

This purpose will make the viewer more comfortable and reduce the eye strain or sickness. (Fig. 5(b) and Fig.5(c))



**Fig. 5.** Result of using gestures and head tracking in 3d stereo anaglyphs: (a) 3d stereoscopy; (b) moving object by using hand gestures; (c) object when head pose moves left;

We implement and run our 3D stereoscopic system on Windows 7 Ultimate version with service pack 1. Our test platform is a CPU 3.7GHz Intel Core i7, NVIDIA GTX 480 graphics card as well as 16GB of RAM and the 22-inch screen. Our system runs fast and smoothly with an average of 60 FPS.

## 5 Conclusion and Future Work

We present a virtual reality system with interactivity controlled by human's view and gestures using the 3D stereoscopic technology which has been popular and used a lot recently. Depending on the current technology of 3D stereoscopic display, our system can only support one viewer and has some limitation in the direction and orientation of the viewer. We demonstrate the effectiveness and the speed of our system in comparison with the related works and approaches by using GPU-based shaders for vertex and fragments. By using KINECT for tracking the gestures and head pose, this is a main idea in our work because it can make the viewer interact with system and enhance the reality when they watch the 3d object in stereoscopic via glasses. Therefore, our real-time stereo rendering system can be extended to the typical and simple virtual reality systems for learning, entertainment, etc.

In the future, we will extend our research with multiple objects displaying parallel in the system and processing hand tracking with depth images calculation. In addition, we will get the value of environment lights, then calculate the ambient, diffuse, and specular light to render anaglyph based 3d stereo objects more realistic.

**Acknowledgment.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0015620).

## References

1. Pollock, B., Burton, M., Kelly, J.W., Gilbert, S., Winer, E.: The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 18(4) (April 2012)

2. Cudalbu, C., Anastasiu, B., Grecu, H., Buzuloiu, V.: Using stereo vision for real-time head-pose and gaze estimation. U.P.B. Sci. Bull., Series C 68(2) (2006)
3. Hirsch, M., Lanman, D.: <http://web.media.mit.edu/~mhirsch/byo3d/>
4. Blondé, L., Doyen, D., Thierry Borel Technicolor Research, Rennes, France: 3D Stereo Rendering Challenges and Techniques. In: 44 th Conference on Information Sciences and Systems, Princeton, March, 17-19 (2010)
5. Frahm, J.-M., Koeser, K., Grest, D., Koch, R.: Markerless Augmented Reality with Light Source Estimation for Direct Illumination
6. Garstka, J., Peters, G.: View-dependent 3D Projection using depth-Image-based Head Tracking. In: 8th IEEE International Workshop on Projector-Camera Systems, Colorado Springs (June 24, 2011)
7. de Sorbier, F., Nozick, V., Biri, V.: GPU rendering for autostereoscopic displays. In: The Fourth International Symposium on 3D Data Processing, Visualization and Transmission, Atlanta, USA (June 2008)
8. OpenNI SDK, <http://www.openni.org>
9. 3GearSystems, <http://www.threegear.com>
10. Paul Bourke, <http://paulbourke.net/stereographics/stereorender>
11. Animesh Mishra, <http://quiescentspark.blogspot.kr>



# Information Management for Multiple Entities in a Remote Sensor Environment

Peter Venero<sup>1</sup>, Allen Rowe<sup>2</sup>, Thomas Carretta<sup>2</sup>, and James Boyer<sup>1</sup>

<sup>1</sup>InfoSciTex, Dayton, Ohio, United States

<sup>2</sup>Supervisory Control and Cognition Branch, 711th Human Performance Wing,  
Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio, United States  
{peter.venero.ctr, allen.rowe, thomas.carretta,  
james.boyer.1.ctr}@wpafb.af.mil

**Abstract.** Current remote piloted aircraft (RPA) operations typically have one sensor operator dedicated to a single sensor, but this may change in the future. To maintain a clear line of sight, the operator must know which sensor to switch to, especially for a moving target. We researched whether using augmented reality and presenting obstruction information helped operators maintain good situational awareness about sensor target relationships. This study had two independent variables: predictive interface (three levels—none, predictive only, and predictive with rays) and interface configuration (two levels—with and without dedicated sensor screens). The results of this study showed that the predictive interface did not increase the operators' performance; however, their performance did increase when we added the dedicated screens.

**Keywords:** augmented reality, sensor management, RPA, control station.

## 1 Introduction

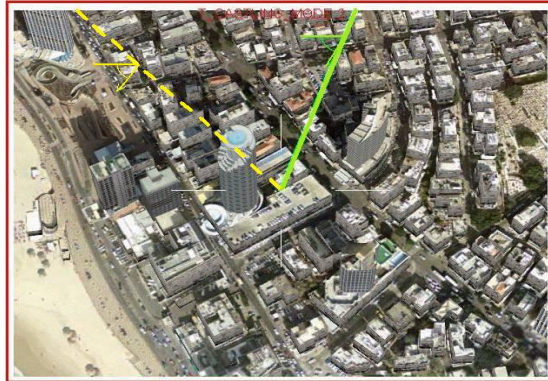
Advances in sensor technology have allowed smaller and smaller remotely piloted aircraft (RPA) to capture full motion video. Ground forces have desired this relatively new capability, and the military expects them to continue to need the technology for some time. To meet this need, the military expects sensor use to increase, making more and more information available to the troops in the field. As the amount of video data increases, the military can conduct more comprehensive and efficient surveillance, but this will require the operator to make more decisions.

In an effort to better understand the impact of multiple RPAs, we chose a persistent stare mission as the experimental task. The persistent stare mission requires the participant to maintain a sensor line of sight on a designated target at all times. One of the main challenges with this task is that terrain features sometimes occlude the sensor view, especially in urban environments. One way to overcome this problem is to fly high and stay directly overhead of the target. Unfortunately, smaller RPAs do not fly very high and are slow, so a fast moving target (such as a car) could out run the RPA. Also, the lower-flying RPA is more likely to be detected.

One strategy to maintain constant eyes on a target with smaller RPAs would be to employ multiple RPAs around an area. Most likely no one RPA will have a clear line of sight to the target at all times, but at least one of them should. The key is for the operator to understand which RPA has the clear line of sight. The purpose of this line of experiments is to investigate interface technology that can help operators gain this understanding.

## 2 Background

We are not the first to use augmented reality and obstruction information. Israeli researchers (Porat, 2010) have previously experimented with using augmented reality. They used augmented reality to help two people maintain persistent stare of a moving target in an urban environment. Figure 1 shows an example of the ‘Castling Rays’ developed by Porat. The Castling Rays provided an operator with information about the elevation of the sensor, the RPA affiliation, and obstruction information (whether or not the sensor could see the desired stare point). The results from Porat showed that these measures increased operators’ performance and situation awareness. The researchers at the Air Force Research Laboratory (AFRL) wanted to follow up the concept of using augmented reality for aiding in decision making. In particular, the AFRL researchers wanted to know what information to encode in the rays.



**Fig. 1.** Sample of the ‘Castling Rays’ augmented reality

The AFRL research consisted of three studies that investigated different aspects of line-of-sight rays and target conditions. In the first study, AFRL required participants to monitor the front door of a building in an urban environment and indicate when persons of interest entered the building. Then, in phase two of the study, a person of interest (POI) leaves a building and participants were to maintain a clear line of sight to the moving target. For the first study, we auto-tracked the moving target. The task for the second and third studies focused only on the moving target task. We told the participants that they were providing support to a customer who was behind enemy

lines and had a small screen with only a single view (like a smart phone or tablet). The customer was relying on them (the participant) to provide constant line of sight of the target. All four sensors would move together, so if the participant changed the stare point, the stare points for the other sensors would move as well. The interface used for all the studies consisted of a map, master sensor screen (the external customer's screen), and multiple dedicated sensor screens.

In the first study (Rowe, 2012), we varied four different aspects of the rays. The ray thickness varied with distance or sensor resolution. We showed obstruction information by using dashed rays when the target was obstructed. For sensor identification, we assigned each RPA a different color. We also varied whether or not the rays were selectable. The results of the first study indicated that the optimal condition for both the stationary and moving target conditions was solid rays that showed sensor identification and were selectable. Contrary to what we would have thought, operators did not perform better when the rays showed obstruction information. We think this outcome was mainly due to the auto slewing of the sensors.

The second study had two independent variables: ray configuration (three levels—rays with obstruction information, rays without obstruction information, and no rays) and interface configuration (two levels—with and without dedicated sensor screens). The results again showed that operators performed better when we did not provide obstruction information on the rays. The dedicated sensor views did not significantly increase operators' performance, but they did reduce operators' workload and were more desirable to the operators.

From the first two studies, it seemed clear that the rays were useful and did improve operators' performance. The dedicated sensor screens also were valuable because they reduced operators' workload. Following these studies, we had several questions that we wanted to investigate in the next study: 'Does adding more sensors make the rays less useful?', 'Can we display the obstruction information in a different way to make it more useful?', and 'Will the dedicated screens increase operators' performance with an increased number of RPAs?'

The third study again had two independent variables: predictive interface (three levels—none, predictive only, and predictive with rays) and interface configuration (two levels—with and without dedicated sensor screens). The results of this study showed that operators did not perform better with the predictive interface; however, they did perform better when we added the dedicated screens.

## 2.1 Hypotheses

The first hypothesis for this paper (h1) is that conditions with the timeline will outperform the conditions without it. The second hypothesis (h2) is that there will be an interaction between the interface configuration and presence of the timeline, meaning the timeline will replace the need for the dedicated screens. Finally, the third hypothesis (h3) is that the conditions that provide obstruction information will outperform the conditions that don't.

### 3 Methods

#### 3.1 Participants

We used a total of 24 (21 males, 3 females) participants for the study. All of the participants were subject matter experts (SMEs) and had experience with using or exploiting remote sensors. Twelve of the participants were from the Springfield Air National Guard (SPANG), and the other twelve participants were from distributed ground station -Indiana (DGS-IN).

#### 3.2 Apparatus

This study used eight computers running the Windows 7 operating system, and all computers were connected to a local network. Six computers ran individual instances of a virtual environment that we developed in-house (Subr Scene). Subr Scene rendered a digital representation of Sadr City in Iraq. We used these computers as sensor feeds for the Vigilant Spirit Control Station (VSCS) (Rowe & Davis 2009) and had their desktops duplicated and streamed digitally for VSCS to consume. We used FFmpeg for both decoding and encoding the video stream in H.264. A fifth computer ran the Vigilant Spirit Simulation and Vigilant Spirit's Zippo. Zippo controls Ternion's Flames Simulation Framework, which we used to send distributed interaction simulation (DIS) packets to control the four computers running Subr Scene. Each participant used a sixth computer running the VSCS in conjunction with two, 24-inch monitors set to a resolution of 1920x1200 pixel, and a Dell laser, 6 button, wired mouse. We permitted each user to adjust their distance from the monitors as desired. Figure 2 shows the participant control station.



**Fig. 2.** Desktop system that participants used

### 3.3 Task

Similar to what we had done in previous studies, we told the participants to provide the best possible sensor feed to a notional joint terminal area controller (JTAC) who was equipped with a small hand-held computer (similar to an Apple iPad) that can only consume one video at a time. In this scenario, the JTAC was relying on the participant to maintain a constant view of a high value individual (HVI). The HVI moved through an urban environment for approximately eight minutes, stopping twice per trial to converse with associates. The trial ended when the HVI entered a building. We randomly placed six RPAs in one of six predetermined starting positions, and they flew a predetermined route. Due the path of the target and the route of the RPAs, no one RPA had a clear line of sight to the target at all times.

### 3.4 Procedures

We gave the participants an introduction to the program, and they reviewed and signed the informed consent form. They then completed three training trials that we had designed to familiarize them with the task and the different aspects of the control station. After they had completed the training trials, they completed 12 data collection trials. After each trial, we administered questionnaires that assessed workload and situation awareness.

### 3.5 Independent Variables

The experiment had two independent variables: obstruction information presentation (OIP) and interface configuration. The OIP independent variable describes how we presented the obstruction information to the participant. The OIP independent variable had three levels: none (no obstruction information presented), timeline only (obstruction information presented through the timeline only), and timeline with rays (obstruction information presented through timeline and the vantage rays). The interface configuration independent variable determined whether or not the operator had dedicated screens for each of the sensors. Figure 3 shows an example of the map (A), master sensor screen (B), timeline (C), and the six dedicated sensor screens (D).

### 3.6 Dependent Variables

We collected objective and subjective measures to assess the participants' performance. We collected the following objective measures: percent unoccluded, average unoccluded time span, total number of transitions, and number of good transitions. The percent unoccluded was the amount of time the HVI was actually visible in the master sensor screen divided by the amount of time the HVI was potentially able to be seen in the master sensor screen. The average unoccluded time span was the average length of time the HVI was in the master sensor screen. The total number of transitions was a count of the number of times the participant switched sensors in the master sensor screen. The number of good transitions was a count of the number of times the participant switched sensors and was able to see the HVI in the new sensor. We also collected subjective measures for the operators' situation awareness and workload.



**Fig. 3.** Sample interface with map (A), master sensor screen (B), timeline (C), and six dedicated sensor screens (D)

## 4 Results

We observed significant differences in only two of the objective measures. Neither of the subjective measures showed any significant differences. For *transition count*, we observed significant effects ( $p < .001$ ) for ray condition (no rays: 16.59; timeline only: 21.86; timeline with rays: 25.44) and interface configuration (master sensor screen only: 24.46; master sensor screen with six dedicated screens: 18.12). A significant interaction ( $p < .05$ ) existed between ray condition and interface configuration when the difference between interface configuration levels was smaller under the no ray condition.

For *user percent unoccluded*, there was a significant main effect ( $p < .05$ ) for interface configuration (master sensor screen only: 84.5%; master sensor screen with six dedicated screens: 86.7%) and an interaction between ray condition and interface configuration.

## 5 Discussion

The underlying belief behind the hypotheses was that the obstruction information would be beneficial; however, the data does not support that theory. When considering the performance metric *user percent unoccluded*, the main factor that improved operators' performance was the dedicated sensor screens. This data does not support hypothesis h1 or h3 (obstruction information would improve operators' performance). Hypothesis h2 postulated that the timeline would remove or reduce the need for the dedicated screen, but we did not see that at all. When looking at the transition counts, which may be an indication of the operators' workload, we see no benefit of providing the participants with the obstruction information; in fact we see an opposite effect. The question is 'Why did we see this?'

We think that the main reason providing obstruction information did not help operators is that the database for our environment contained noisy information. In order to determine if a sensor is being blocked or not, the operator needs a detailed database of the environment. In our experiments we had such a database; however, it did not

account for architectural details (e.g., awnings) or other attributes in the environment such as trees or vehicles. These minor omissions from the database would cause the obstruction information driving the timeline to be incorrect, which would in turn lead the participants to not trust the timeline.

## 6 Conclusion

In conclusion, the data from previous and current research suggests augmented reality rays improve operators' performance when maintaining persistent stare with multiple sensors. One caveat to this research is that all of the tasks took place in an urban environment with straight and either parallel or perpendicular streets, possibly helping the operators in unforeseen ways. The number of RPAs an operator is controlling also influences the utility of augmented reality. In addition, the data suggests that providing obstruction information is not beneficial. Future researchers will examine the appropriate size for the dedicated sensor screens and begin to look at how to employ this technology on mobile devices.

## References

1. Porat, T., Oron-Gilad, T., Silbiger, J., Rottem-Hovev, M.: Castling Rays' a Decision Support Tool for UAV-Switching Tasks. Paper Presented at CHI, Atlanta, GA (April 2010)
2. Rowe, A.J., Davis, J.E.: Vigilant spirit control station: A research test bed for multi-UAS supervisory control interfaces. In: Proceedings of the Fifteenth International Symposium on Aviation Psychology, Dayton, OH, pp. 287–292 (2009)
3. Rowe, A., Venero, P., Boyer, J.: Improving Situation Awareness Through Enhanced Knowledge of Sensor Target Relationships. Paper Presented at Military Operations Society Symposium, Colorado Springs, CO (June 2012)
4. Venero, P., Rowe, A., Boyer, J.: Using Augmented Reality to Help Maintain Persistent Stare of a Moving Target in an Urban Environment. Paper Presented at HFES, Boston, MA (October 2012)

## **Part II**

# **Interaction in Augmented and Virtual Environments**



# Tactile Apparent Motion Presented from Seat Pan Facilitates Racing Experience

Tomohiro Amemiya<sup>1</sup>, Koichi Hirota<sup>2</sup>, and Yasushi Ikei<sup>3</sup>

<sup>1</sup> NTT Communication Science Laboratories,  
3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198 Japan  
[amemiya.tomohiro@lab.ntt.co.jp](mailto:amemiya.tomohiro@lab.ntt.co.jp)

<http://www.br1.ntt.co.jp/people/t-amemiya/>

<sup>2</sup> Graduate School of Interfaculty Initiative in Information Studies,  
The University of Tokyo,

5-1-5 Kashiwanoha, Kashiwano-shi, Chiba 277-8563 Japan

<sup>3</sup> Graduate School of System Design, Tokyo Metropolitan University,  
6-6 Asahigaoka, Hino-shi, Tokyo 191-0065 Japan

**Abstract.** When moving through the world, humans receive a variety of sensory cues involved in self-motion. In this study, we clarified whether a tactile flow created by a matrix of vibrators in a seat pan simultaneously presented with a car-racing computer game enhances the perceived forward velocity of self-motion. The experimental results show that the forward velocity of self-motion is significantly overestimated for rapid tactile flows and underestimated for slow ones, compared with only optical flow or non-motion vibrotactile stimulation conditions.

**Keywords:** Tactile flow, Optic flow, Multisensory integration.

## 1 Introduction

In chair-like vehicles, humans generally detect velocity information using visual cues and detect acceleration and angular acceleration information using mechanical cues (vestibular and tactile sensations). The authors have been focusing on developing multisensory displays, such as vestibular and tactile displays, to facilitate self-motion perception in a vehicle-based VR system [2,18,3], which present multiple modality stimuli to produce a subjective realistic experience [12].

A stationary observer often feels subjective movement of the body when viewing a visual motion simulating a retinal optical flow generated by body movement [22]. Gibson concluded that a radial pattern of optical flow is sufficient for perceiving the translational direction of self-motion [9]. In virtual reality environments, not accurate speed but differences in locomotion speed are perceived from an optical flow [4,5]. The self-motion generated by the optic flow is further facilitated by adding a sound moving around the user [17], a constant flow of air to the user's face [19], or simple vibrotactile cues from a seat [16,7]. Even when tactile and visual stimuli indicate motions in different directions, sensitivity to motion can be improved [10]. When consistent somatosensory cues are added to

the hand (a sustained leftward pressure on a fist for rightward visual rotation), the cues also facilitate vection [15].

The conventional tactile seat vibration approach has been used for directing visual attention or displaying directional information [21,11,13]. For example, Tan et al. developed a  $3\times 3$  tactile matrix display built into a back rest or seat to provide directional or way-finding information [21]. Hogema et al. conducted a field study with an  $8\times 8$  matrix of tactors embedded in the seat pan in car to indicate eight different directions [11]. Israr and Poupyrev proposed an algorithm to create a smooth, two-dimensional tactile motion for a  $4\times 3$  tactile matrix display and applied it to computer games [13]. Our approach is different from these previous studies in that we focus on the tactile seat vibration approach to create a forward velocity change of self-motion.

The actuators in most tactile matrix displays are arranged sparsely. However, the apparent motion, the illusory perception of motion created by the discrete stimulation of points appropriately separated in space and time, can be experienced between pairs of touches, as well as pairs of lights or sounds. With the stimuli generating optimum apparent motion, the user would not perceive two discrete tactile stimuli but rather a single moving tactile stimulus between the two, regardless that the tactile stimulus is not actually moving on the surface of the skin [6,8]. The tactile apparent motion is elicited with two main parameters: stimulus duration and the inter-stimulus onset interval (ISOI, often referred as stimulus onset asynchrony: SOA [14]) between onsets of subsequent tactile stimuli [20]. Both parameters determine the velocity of illusory movement. In our study, we used seat vibration to provide a velocity cue of self-motion, which was varied by changing the ISOI between the onsets of sequentially activated rows of vibrators.

## 2 Experiment: Tactile Flow with Car Racing Video Game

To test the feasibility of enhancing perceived forward velocity of self-motion in a virtual environment, we conducted an experiment using ten-second videos of a car-racing computer game with our tactile feedback device. We also evaluated the perception of moving forward velocity using an expanding radial flow motion in peripheral vision and a tactile flow from a seat pan. In both experiments, participants viewed optical stimuli while gazing at a fixation cross and sitting on a tactile stimulator on the seat pan and rated their perceived forward velocity by using the method of magnitude estimation.

### 2.1 Method

Six participants (three males and three females; 18-35 years old) participated. Recruitment of the participants and the experimental procedures were approved by the NTT Communication Science Laboratories Research Ethics Committee, and the procedures were conducted in accordance with the Declaration of Helsinki. None of the participants was aware of the purpose of the experiment.

The tactile stimulator is composed of twenty voice-coil motors arranged in a  $4 \times 5$  grid and an aluminium plate (196-mm long  $\times$  245-mm wide  $\times$  3-mm thick, A5052P) with holes. The voice-coil motor is a full-range speaker (NSW2-326-8A; Aura Sound Inc.), and presents stronger stimuli in a wider area than conventional eccentric rotating mass vibration motors. Each voice-coil motor was attached to pin-arrays made of ABS resin and vibrated sinusoidally at 50 Hz for 200 ms of stimulus duration. The pins popped up from the holes and vibrated vertically. Five voice-coil motors in the same lateral line were driven together by a computer with D/A boards (DA12-16(PCI) and DA12-8(PCI); CONTEC Co., Ltd.) and a custom-made circuit (including an amplifier). They were sequentially activated with a constant interval, which created a tactile motion from front to back. We varied the ISOI (100, 200, or 300 ms) to change the velocity of the tactile flow. In the conditions where the ISOIs were greater than the stimulus duration (200 ms), there was a blank interval between tactile stimuli. In addition, we used two control conditions for tactile flow: a non-motion random vibration (five vibrators out of twenty randomly and successively activated with a 200-ms duration and 200-ms ISOI); no vibration (i.e., vision-only condition).

Participants were seated on the center of the tactile device. One side of the tactile device was parallel to the monitor. They were instructed to keep their heads on a chin rest as shown in Fig. 1. The participants observed the stimulus binocularly. Participants wore an earmuff (Peltor Optime II Ear Defenders; 3M, Minnesota, USA) to mask the sound of the tactile device. Subjects were instructed to watch the fixation cross shown at the center of the stimulus display during the trial. The standard stimuli consisting of visual motion were presented for ten seconds. After a two-second pause, the test stimuli consisting of visual and/or tactile motion were presented for ten seconds. During the experiment, the experimental room was darkened. No lighting other than the monitor was present in the room.

A recorded video of a car-racing computer game (TORCS; The Open Racing Car Simulator [1]) was presented as a visual stimulus. The video gives a first-person perspective while driving the car on a straight stretch of a racing track. The playback speeds of the video were changed (133%, 100%, or 66%), and then the length of the videos were set to identical lengths (ten seconds). A fixation cross was always shown at the center of the video, and the participant was instructed to gaze at it during the trial. In the video, there were some cues related to optical flow, such as road texture, signboards on a wall, and some buildings. Figure 2 shows the temporal sequence of the experiment. The participants' task was to estimate the velocity of self-motion. Two trials were conducted for each condition. The presentation order of the fifteen conditions (five tactile conditions  $\times$  three video playback speeds) was pseudo-randomized.

Visual stimuli of control condition were radial expansions of approximately 1,000 random dots in each frame, simulating a translational motion. The dots had a diameter changing from six pixels (corresponding to 0.5 degrees) to thirty pixels (2.5 degrees) according to the distance from the center. The visual stimulus images (1,024  $\times$  768 pixel resolution at 60-Hz refresh rate) were presented on

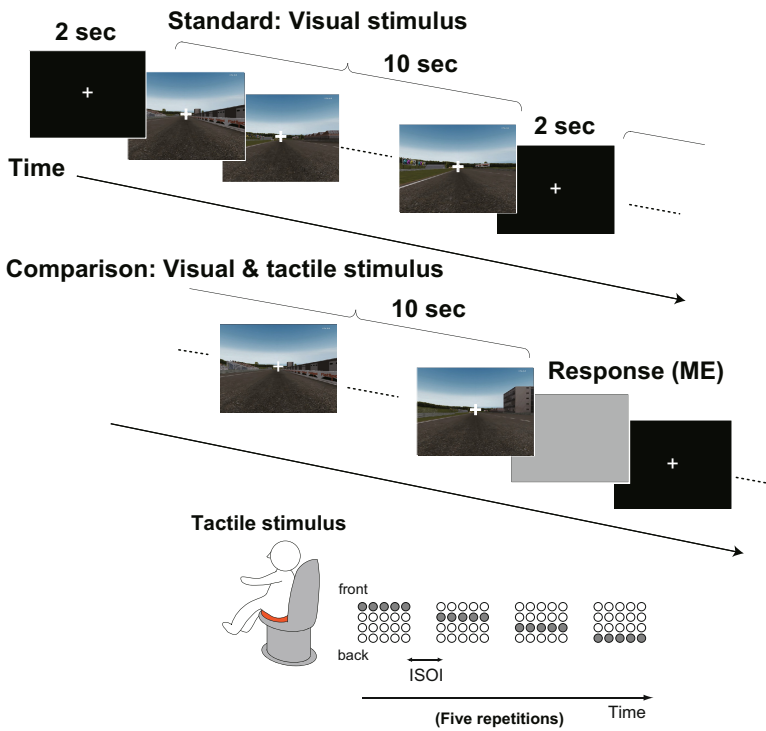


**Fig. 1.** Configuration of the experimental apparatus. The subject was instructed to sit on the tactile stimulator and gaze at the fixation cross at the center of the monitor. The distance between the monitor screen and the subject's face was 30 cm. Experiments were performed in a darkened room.

a 21.3-inch LCD screen (Iiyama Inc.). The images subtended a visual angle of 72 deg (horizontal)  $\times$  57 deg (vertical) at the viewing distance of 300 mm. The central circular area of the diameter of 20 degrees was masked with a black background in such a way that the moving dots were presented only outside the circular border. The velocity of the expanding optical flow was changed from 80% to 120% of the standard stimulus.

## 2.2 Results and Discussion

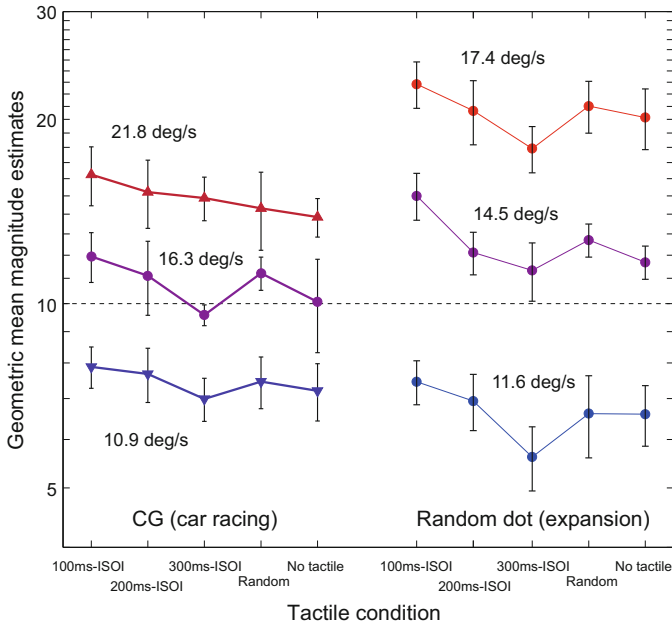
Figure 3 shows the averaged values (geometric mean) obtained by magnitude estimation. The perceived velocity of self-motion was facilitated by both the velocities of the tactile motion and the video playback, which is consistent with the result of the control condition using the random dot pattern. At 133% video playback speed, the perceived velocity of self-motion increased as the ISOI of the tactile flow decreased (i.e., increased with the speed of the tactile flow). In addition, the perceived velocity under the random vibration condition was



**Fig. 2.** Temporal sequence. A 10-sec video of a car-racing computer game was used as a visual stimulus. The video playback speed was altered. The tactile stimulus was consisted of four successive rows of vibration. The inter-stimulus onset interval (ISOI) between the tactile rows was varied to measure the effect on the perceived velocity of self-motion. Participants had to report numerically their perceived velocity relative to the standard in a magnitude estimation.

smaller than others at 133% video playback speed, perhaps because the tactile stimuli of random vibration were not consistent with self-motion. On the other hand, at 66%, there was not any clear difference across the tactile conditions.

To quantitatively evaluate the difference between the conditions, we conducted a two-way repeated measures ANOVA for magnitude estimation values. The analysis revealed significant main effects of the video playback speed condition [ $F(2,10) = 33.68, p < .001, \eta_p^2 = .87$ ]. But the main effect of the tactile condition was not found [ $F(4,20) = 2.01, p = .13, \eta_p^2 = .29, n.s.$ ]. There was no significant interaction between the video playback speed condition and tactile condition [ $F(8,40) = 0.27, p = .97, \eta_p^2 = .05, n.s.$ ]. These results indicate that tactile apparent motion on a seat pan did not facilitate the perceived forward velocity of self-motion in a racing game application as much as it did with expanding random dot pattern stimuli. Future work includes conducting further experiments with more participants to show the effect more correctly.



**Fig. 3.** Magnitude estimation for forward velocity as a function of ISOI of the tactile stimuli (left: CG, right: random dot pattern). Each dot represents the geometric mean value across participants. Error bars show SEs.

### 3 Conclusion

In this paper, we have shown experimentally a change in perceived forward velocity of self-motion caused by changes in the speeds of the tactile apparent motion on a seat pan. When visual and tactile flows that simulate forward moving are presented simultaneously, the quicker tactile motion stimuli enhanced the perceived forward velocity of self-motion and the slower ones inhibited it. In contrast, almost no change in velocity perception was observed when a tactile stimulus without motion cues was presented together with an optic flow. Finally, we confirmed that the method using tactile flow on the seat pan can be applied in a car-racing computer game. So far, the method seems to change the perceived velocity a bit but not to function as well as when a simple expanding optic flow is presented.

Future work will investigate whether it is possible to change the perceived velocity by changing the stimulus duration or the intensity of tactile stimuli. We will also examine the effect of the perceived change in moving velocity with contraction of the random dots or backward motion of tactile flow.

**Acknowledgements.** This research was supported by the National Institute of Information and Communication Technology (NICT), Japan.

## References

1. The open racing car simulator website (2012), <http://torcs.sourceforge.net/>
2. Amemiya, T., Hirota, K., Ikei, Y.: Concave-convex surface perception by visuo-vestibular stimuli for five-senses theater. In: Shumaker, R. (ed.) *Virtual and Mixed Reality*, HCII 2011, Part I. LNCS, vol. 6773, pp. 225–233. Springer, Heidelberg (2011)
3. Amemiya, T., Hirota, K., Ikei, Y.: Perceived forward velocity increases with tactile flow on seat pan. In: *Proc. IEEE Virtual Reality Conference*. IEEE Computer Society, Los Alamitos (2013)
4. Banton, T., Stefanucci, J., Durgin, F., Fass, A., Proffitt, D.: The perception of walking speed in a virtual environment. *Presence: Teleoperators and Virtual Environments* 14(4), 394–406 (2005)
5. Bruder, G., Steinicke, F., Wieland, P., Lappe, M.: Tuning self-motion perception in virtual reality with visual illusions. *IEEE Transactions on Visualization and Computer Graphics* 18, 1068–1078 (2012)
6. Burtt, H.E.: Tactual illusions of movement. *Journal of Experimental Psychology* 2, 371–385 (1917)
7. de Vries, S.C., van Erp, J.B., Kiefer, R.J.: Direction coding using a tactile chair. *Applied Ergonomics* 40(3), 477–484 (2009)
8. Geldard, F.A., Sherrick, C.E.: The cutaneous “rabbit”: A perceptual illusion. *Science* 178(4057), 178–179 (1972)
9. Gibson, J.J.: *The Perception of the Visual World*. Houghton Mifflin (1950)
10. Gori, M., Mazzilli, G., Sandini, G., Burr, D.: Cross-sensory facilitation reveals neural interactions between visual and tactile motion in humans. *Frontiers in Psychology* 2(55), 1–9 (2011)
11. Hogema, V.E., De Vries, Kiefer: A tactile seat for direction coding in car driving: Field evaluation. *IEEE Transactions on Haptics* 2(4), 181–188 (2009)
12. Ikei, Y., Abe, K., Hirota, K., Amemiya, T.: A multisensory vr system exploring the ultra-reality. In: *Proc. 18th International Conference on Virtual Systems and Multimedia (VSMM)*. IEEE (2012)
13. Israr, A., Poupyrev, I.: Tactile brush: drawing on skin with a tactile grid display. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2019–2028. ACM (2011)
14. Kirman, J.H.: Tactile apparent movement: The effects of interstimulus onset interval and stimulus duration. *Perception and Psychophysics* 15, 1–6 (1974)
15. Lécuyer, A., Vidal, M., Joly, O., Mégard, C., Berthoz, A.: Can haptic feedback improve the perception of self-motion in virtual reality? In: *Proc. IEEE Haptic Symposium*, pp. 208–215 (2004)
16. Riecke, B.E., Schulte-Pelkum, J., Caniard, F., Bulthoff, H.H.: Towards lean and elegant self-motion simulation in virtual reality: vibrational cues enhance the perception of illusory self-motion. In: *Proc. IEEE Virtual Reality Conference*, pp. 131–138 (2005)
17. Riecke, B.E., Väljamäe, A., Schulte-Pelkum, J.: Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception* 6(2), 7:1–7:27 (2009)
18. Saito, T., Ikei, Y., Amemiya, T., Hirota, K.: Sound and vibration integrated cues for presenting a virtual motion. In: *Proc. International Conference on Augmented Tele-existence (ICAT)*, pp. 216–217 (2010)

19. Seno, T., Ogawa, M., Ito, H., Sunaga, S.: Consistent air flow to the face facilitatesvection. *Perception* 40(10), 1237–1240 (2011)
20. Sherrick, C., Rogers, R.: Apparent haptic movement. *Perception and Psychophysics* 1, 175–180 (1966)
21. Tan, H.Z., Gray, R., Young, J.J., Traylor, R.: A haptic back display for attentional and directional cueing. *Haptics-e* 3(1) (2003)
22. Warren, W.H., Hannon, D.J.: Direction of self-motion is perceived from optical flow. *Nature* 336(6195), 162–163 (1988)



# Predicting Navigation Performance with Psychophysiological Responses to Threat in a Virtual Environment

Christopher G. Courtney<sup>1,2</sup>, Michael E. Dawson<sup>1</sup>, Albert A. Rizzo<sup>2</sup>,  
Brian J. Arizmendi<sup>3</sup>, and Thomas D. Parsons<sup>4</sup>

<sup>1</sup>Department of Psychology, University of Southern California, USA

<sup>2</sup>Institute for Creative Technologies, University of Southern California, USA

<sup>3</sup>Department of Psychology, University of Arizona, USA

<sup>4</sup>Department of Psychology, University of North Texas, USA

courtney@ict.usc.edu, {dawson, arizzo}@usc.edu,  
arizmendi@email.arizona.edu, thomas.parsons@unt.edu

**Abstract.** The present study examined the physiological responses collected during a route-learning and subsequent navigation task in a novel virtual environment. Additionally, participants were subjected to varying levels of environmental threat during the route-learning phase of the experiment to assess the impact of threat on consolidating route and survey knowledge of the directed path through the virtual environment. Physiological response measures were then utilized to develop multiple linear regression (MLR) and artificial neural network (ANN) models for prediction of performance on the navigation task. Comparisons of predictive abilities between the developed models were performed to determine optimal model parameters. The ANN models were determined to better predict navigation performance based on psychophysiological responses gleaned during the initial tour through the city. The selected models were able to predict navigation performance with better than 80% accuracy. Applications of the models toward improved human-computer interaction and psychophysiological-based adaptive systems are discussed.

**Keywords:** Psychophysiology, Threat, Simulation, Navigation, Route-Learning.

## 1 Introduction

The incorporation of simulation technology into neuroergonomic and psychophysiological research is advancing at a steady rate (see [1], [2]). The range and depth of these simulations cover a large domain, from simple low fidelity environments to complex fully immersive simulators, which are factors that may affect psychophysiological response within the environment [3]. All of these simulators rely on some type of representation of the real world [4]. The current study utilized a high fidelity, highly immersive virtual environment, as increased applicability to real-world performance was the goal. Specifically, the virtual

environment (VE) utilized herein was that of a virtual Iraqi city [5], which included a route-learning and navigation simulation to assess landmark and route knowledge of the newly experienced VE [6]. Psychophysiological responses were monitored throughout the experiment and were used to predict navigation performance following threat exposure during the route-learning phase.

## 1.1 Navigation in Virtual Environments

Numerous studies have conveyed the benefits of ecologically valid simulated navigation tasks as predictors of real-world functioning [4], [7], [8]. Navigation abilities are customarily broken down into three knowledge based components, each adding to the cognitive map developed by the participant. The first is landmark knowledge, which involves learning to recognize landmarks or salient features of the environment upon initial exploration of said environment [9]. In the current study, zone markers indicating the entrance into a new zone were the key landmarks involved. The second component is referred to as procedural or route knowledge and involves information gleaned from first-hand experience with a route which provides the ability to create distance and orientation relationships connecting landmarks [9], [10]. Real-world and virtual reality (VR) experiments suggest that active navigation, which was utilized in the current research, is more effective for route-learning than passively being exposed to the environment [11]. The third component of navigation ability is referred to as survey knowledge, which can be described as having developed a “bird’s eye view” of the environment. Survey knowledge affords the development of a cognitive map that provides associations between locations with increased levels of exposure to the environment [9], [12]. Survey knowledge is valuable as a means of finding shortcuts through the environment, but is not necessarily useful in the present study, as participants were instructed to follow a specific route without deviating. Thus, this study is primarily concerned with landmark and route knowledge.

The current research design afforded the opportunity to investigate the effects of exposure to threat on route-learning. To our knowledge, no study has examined the effects of varying levels of threat on route-learning, making this a novel approach. We hypothesized that threatening stimuli in the environment would serve as distractors and would hinder route-learning in highly threatening areas of the VE. Past research involving distractors presented during route-learning typically involve cognitive workload tasks and tend to interfere with route-learning. Walker and Lindsay [13] reported decreased efficiency in wayfinding performance in a virtual city when a secondary speech discrimination task was introduced. They postulate that this was due to the switch of attentional resources to the completion of the secondary task. A similar result was found in a between-subjects study involving examination of the effects three separate types of cognitively distracting tasks presented during the route-learning phase compared to a no task condition. All groups that experienced distracting tasks performed less efficiently on a wayfinding task than the group that was not presented with any distracting task [14]. Knowledge of psychophysiological states gleaned during the route-learning phase may serve as an indicator of

wayfinding abilities. For example, participants with lower psychophysiological response levels during the route-learning phase may prove more efficient during the navigation phase.

## **1.2 Toward Adaptive Simulations**

The current research was concerned with informing psychophysiological computing strategies for creation of VEs capable of adapting to the participant's affective and cognitive state to foster optimal performance. Psychophysiological computing represents an innovative mode of human-computer interaction (HCI) wherein system interaction is achieved by monitoring, analyzing and responding to covert psychophysiological activity from the user in real-time [15], [16]. Psychophysiological computing represents a means of creating for the computer system a more empathic link to the user.

The strategy employed herein for creation of a psychophysiological computing system initially required assessment of psychophysiological response patterns associated with varying affective states. The current research manipulated environmental threat to create response variability in order to perform such assessments. Data analytic approaches designed for prediction were then compared and tested for effective development of a psychophysiological computing system capable of predicting performance outcomes. Namely, the efficacy of multiple linear regression (MLR) and artificial neural network (ANN) models were compared for prediction of navigation performance based on psychophysiological responses to threat and cognitive workload during the route-learning phase of the experiment.

In summary, participants were exposed to varying levels of threat while concurrently completing a route-learning task, and the responses collected were submitted to MLR and ANN models to predict performance on the subsequent test of route-learning efficacy during a navigation task.

## **2 Methods**

### **2.1 Participants**

A total of 53 participants (67.9% female; mean age = 19.79; age range = 18 to 22) took part in the experiment. Participants were recruited through the psychology subject pool at the University of Southern California. Inclusion criteria included normal or corrected to normal vision, and English fluency. Participants were between the ages of 18 and 35.

### **2.2 Stimuli**

A virtual environment depicting an Iraqi city was presented to participants with use of an eMagin Z800 head mounted display complete with head tracking capabilities to allow the participant to explore the environment freely. The virtual environment was created using graphic assets from the virtual reality cognitive performance test

(VRCPAT) [6], [17], using the Gamebryo graphics engine to create the environment. A tactile transducer floor was utilized to enhance the ecological validity of the VE by making explosions and other high threat stimuli feel more lifelike. Auditory stimuli were presented with a Logitech surround sound system. Psychophysiological measures related to electrodermal and electroencephalographic activity were collected using a Biopac MP150 system. Participants experienced the VE while residing in an acoustic dampening chamber, which had the added benefit of creating a dark environment to remove any peripheral visual stimuli that were not associated with the VE, resulting in increased immersive qualities of the simulation.

### **2.3 Procedural Design**

Following a baseline procedure, participants were exposed to the route-learning task. The task consisted of following a guide through six zones that alternated between high and low levels of threat. All environmental stimuli were pre-scripted, allowing each participant to experience exactly the same environmental stimuli at the same time to enhance experimental control of stimulus presentation. The high and low zone presentation order was counterbalanced across subjects as to which type of zone was experienced first. During the high threat zones, participants experienced an ambush situation in which bombs, gunfire, screams and other visual and auditory forms of threat were present, whereas none of these stimuli were presented in the low threat zone. Each zone was preceded by a zone marker, which served as landmarks to assist in remembering the route.

The route-learning task was followed immediately by the navigation task in which the participants were asked to return to the starting point of their tour through the city. Participants were to pass through each zone in reverse order until reaching the original starting point. If the participant strayed too far from the path, which was quantified as the distance it would take to walk for 10 seconds in a perpendicular direction from the original path, an arrow appeared in the corner of the screen that assisted the participant in finding his or her way back to the original path. During the navigation task, there were no longer any threatening stimuli presented in the high threat zones. The navigation task ended when the participant crossed the zone 1 marker.

### **2.4 Analytic Approach**

Data were scored using an in-house custom designed Matlab scoring program. The program includes graphical representations of each channel of psychophysiological data for manual inspection of scoring accuracy.

**Electrodermal Data Scoring.** The scoring program was used to partition response levels into each zone, and then calculate the median skin conductance level (SCL) and the number of spontaneous fluctuations (SFs) in each. The median SCL was chosen for analyses rather than the mean because it is a more robust feature as it is less susceptible to influences of artifacts, which will be especially useful in future adaptive

applications. SFs, which were also scored during trimmed zones, were quantified as any change in slope of the response curve resulting in a  $> 0.01 \mu\text{S}$  response, with a peak latency of 1 to 3 seconds following onset.

**Electrocardiographic Data Scoring.** ECG data were scored as inter-beat intervals (IBIs), which were calculated as median values for each zone. Accuracy of the peak detection scoring program was assessed manually, with visual inspection of all selected R-waves. Missed R-waves were manually added to the calculation of zone medians. Power spectral density analyses of heart rate variability (HRV) were also performed with use of a fast Fourier Transform based algorithm. The algorithm was used to calculate the spectral power of the low frequency (LF) component and the high frequency (HF) component of HRV associated with each zone. The frequency range of the LF component is between 0.04 and 0.15 Hz, while the HF component is between 0.15 and 0.4 Hz [18].

**Respiratory Data Scoring.** Respiration was scored in a similar fashion to the ECG data, and reported as interbreath intervals. Peak detection of each positive deflecting curve in the breathing cycle was manually reviewed in order to ensure accuracy of the scoring program, and median intervals were calculated for each zone.

**MLR and ANN Approach.** The experimental conditions described herein are designed to provoke responses typical of high and low extremes of experienced threat. The ultimate purpose of the proposed ANNs is to develop a strategy for creating adaptive systems for future research and eventual real-world applications including enhanced training scenarios and adaptive assistance for any individual who must fulfill tasks that involve high levels of threat, stress, or cognitive effort. A backpropagated algorithm was utilized to train the ANN models mainly because it can be thought of as a specialized case of the general linear model that is capable of more effectively fitting curvilinear data distributions than is possible with a linear regression model. Additionally, because the ANN model can be thought of as a special type of regression, and provides similar output, results can be compared directly to predictive results generated with the use of more standard and widely used MLR. This sets the backpropagated algorithm apart from numerous machine learning algorithms, such as support vector machines, which can lead to difficulties when trying to compare causes for predictive differences with other algorithms.

First, a MLR model that used the psychophysiological data gathered during the initial tour through the city to predict the navigation performance was developed. The navigation performance was quantified as the time needed to return to the starting point. A set of six psychophysiological predictors were utilized. Included in the analyses were SCLs, SFs, IBIs, interbreath intervals, and the LF and HF components of the HRV measure. Due to the relatively small sample size in this experiment, an attempt to condense the number of predictors was made by calculating difference scores between the high and low threat zones for each of the psychophysiological predictors. Difference scores were calculated in two ways. First, the overall difference

between all three high and low threat zones was calculated as a representation of the response levels associated with the task as a whole. Next, a difference score that would serve as an index of the habituation involved in the responses to the high threat zones compared to the low threat zones was calculated. To accomplish this, difference scores between the high and low threat zones in the first pair of zones experienced in the route-learning phase (zone pair A) and the third pair (zone pair C) were calculated. The zone pair C difference score was then subtracted from the zone pair A difference score. This threat habituation index was calculated to account for the waning response levels during high threat zones present in a number of response measures. Analyses not reported here determined that the habituation-sensitive predictors were preferable. A backward-elimination stepwise regression was utilized for the MLR model.

The ANN model was developed in a manner analogous to the above MLR model, such that the predictor variables, or inputs, were the same in each model. The output node will again represent the continuous navigation performance outcome measure of the time needed to return to the starting point. The primary goal of the BPN used herein is prediction. In order to increase the probability of generalization and to avoid over-fitting of the observed sample, three data sets were considered, including the training set, validation set, and the test set (see [18] for review of these data sets). The test set contains a set of examples that had not been previously considered during the training or validations phases, which is used to calculate the global predictive ability of the network for generalizations to future practical applications. After the development and implementation of the ANN, comparisons were made (following [18]) between its output and that of the general linear model's regression for the predicted outcome measure.

## 3 Results

### 3.1 Regression Results

The MLR model was able to explain a significant proportion of the variance in navigation performance,  $R^2 = 0.27$ ,  $F(7, 45) = 2.32$ ,  $p < 0.05$ . Significant predictors included SFs,  $\beta = -0.34$ ,  $t = 2.66$ ,  $p < 0.05$ , and interbreath intervals,  $\beta = 0.28$ ,  $t = 2.13$ ,  $p < 0.05$ . The negative correlation coefficient related to the SF measure indicates that participants who had a greater difference between the high and low threat zones during zone pair A than zone pair C, due to habituation in high threat zones, took less time navigating back. Though interbreath intervals correlation coefficient was positive, the results are analogous to those of the SFs. Increased activation leads to more SFs and shorter interbreath intervals, so the response patterns are reversed. Thus, greater reduction in differential activation between high and low threat zones during zone pair C resulted in more efficient navigation performance.

**Table 1.** MLR model summary statistics

RMSE = root mean squared error.

<i>R</i>	<i>R</i> <sup>2</sup>	Adj. <i>R</i> <sup>2</sup>	Std. Error	RMSE	<i>F</i>	<i>P</i>
0.52	0.27	0.17	48538.0	220.3	2.32	<0.05

### 3.2 ANN Results

The backpropagated ANN that was developed included the same six predictor variables used in the preferred MLR model, here entered as inputs to the system. In the preliminary tests to assure that the ANN achieved its optimal output, the network model was developed with different numbers of nodes in the single hidden layer. The hidden layer learns to provide a representation for the inputs through an alteration of the weights associated with each node and then connects to the output layer. The experimental method involved developing a hidden layer that contained a minimum of four nodes and a maximum of twenty-four nodes. It was found that six hidden layer nodes resulted in optimal model performance. A tanh activation function was applied to the hidden and output nodes, which is recommended when the sum of squares error function is employed, as it was in this case. Descriptive statistics associated with the training, validation, and test set samples are included in Table 3.

Following network training, the test set was applied to the network to test the generalizability of the model. It should be noted that the predictor values of the test set were not involved in the training of the model, providing a “test” of the generalizability of the model to new data. A gradient descent learning algorithm was applied along with a sum of squares error function. Hyperbolic tangent activation functions were applied to the hidden and output nodes. The ANN was able to predict the outcome measure with 76.0% accuracy (training performance = 0.938; test set performance = 0.871).

A global sensitivity analysis was performed in order to determine the relative importance of each input (i.e., predictor variable) to the successful prediction of the output. A sensitivity analysis tests how the error rates would increase or decrease if each individual input value were changed (see [20] for review). More specifically, the data set is repeatedly submitted to the network, and in turn each input variable is replaced with its mean value calculated from the training sample, and the resulting network error is recorded. Important inputs cause for a large increase in error, while the error increase was small for unimportant inputs. Thus, sensitivity analysis allows for a rank order of the importance of the individual inputs [21, 22]. Ratio values less than 1 indicate that the network actual performs better without inclusion of the associated input. All inputs had ratio values of greater than 1, indicating that all contributed to the performance of the model. The highest ranked inputs were SCLs, IBIs, SFs, and interbreath intervals, each having a ratio value greater than 4.

### 3.3 ANN and MLR Comparisons

Examination of the squared correlation coefficient associated with each model reveals that there is a 49.0% increase in prediction of navigation performance when the ANN is employed. The drop in root mean squared error related to use of the ANN (RMSE = 205.39) in comparison the MLR model (RMSE = 220.30) signifies that the neural network model better fits the data. Direct comparison of correlation coefficients associated with each model with use of the Fisher  $z$  transformation revealed that the ANN had significantly greater predictive ability than the MLR model,  $z = 3.84$ ,  $p < 0.001$ . Thus, the ANN was determined to be the preferable model due to the increase in the squared correlation coefficient in addition to the decrease in RMSE.

## 4 Discussion

The current research offers a number of beneficial design advances for potential use in future training simulation technologies and adaptive systems in general. A VE was developed that was capable of providing a route-learning scenario and the ability to test route-knowledge with use of a navigation task. Manipulations embedded within the VE also afford the opportunity to test the effects of varying levels of threat in the environment. Models were designed to predict navigation performance based on psychophysiological response measures collected during the route-learning phase. Evidence presented led to the conclusion that ANNs were better able to predict performance outcomes, and were generalizable to previously unseen data following training of the model. The goal of this study was to develop strategies for the successful development of systems that utilize psychophysiological computing to adapt to the individual in such a way that an optimal pace for training is achieved in order to foster ideal learning settings. A number of findings reported in the current research provide informative material for such adaptive system development.

Adaptive automation systems generally utilize psychophysiological responses to assess user-states in order to determine the necessity of automated assistance to facilitate optimal system performance [1]. In the current study, habituation effects on threat responses led to the calculation of predictor variables better suited for navigation performance prediction. Responses to threat habituated almost universally throughout the task. Thus, a set of predictors designed to account for habituation effects produced better prediction of navigation performance. This distinction could be used to inform future adaptive system design in that thresholds for adaptations based on responses to threatening stimuli must be concerned with the change in response levels with repeated exposure to the stimuli and must allow for dynamic adjustment to thresholds for adaptive change.

Finally, the current research provided encouraging support for the use of ANNs for prediction of performance outcomes based on psychophysiological response measures. The ANN provided significantly enhanced predictive abilities compared to a traditional MLR model. This demonstrated that psychophysiological responses to varying levels of threat during a route-learning task could be used to predict performance on a subsequent navigation task with better than 76% rates of accuracy.



Recently, researchers have begun applying advanced algorithms such as ANNs for data classification in real-time. For example, a number of studies have utilized ANNs for the initiation of adaptive assistance when features meet classification requirements for a state of overload [23, 24]. These techniques are often used for assessment and classification of nonlinear data (see [19]). The models produced in the current research lend themselves well to use in adaptive training simulations to enhance route-learning abilities when confronted with threatening stimuli. An adaptive automation approach can be employed to training making use of the VE developed herein, such that psychophysiological responses gleaned during the route-learning phase can be assessed for hyper- or sub-threshold criteria related to overload or fear, and adaptive assistance may be provided during the navigation task to fit the needs of the individual and promote optimal performance.

## References

1. Parasuraman, R., Wilson, G.F.: Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors* 50, 468–474 (2008)
2. Parsons, T.D., Courtney, C.: Neurocognitive and Psychophysiological Interfaces for Adaptive Virtual Environments. In: Röcker, C., Ziefle, M. (eds.) *Human Centered Design of E-Health Technologies*, pp. 208–233. IGI Global, Hershey (2011)
3. Parsons, T.D., Rizzo, A.A., Courtney, C., Dawson, M.: Psychophysiology to Assess Impact of Varying Levels of Simulation Fidelity in a Threat Environment. *Advances in Human-Computer Interaction* 5, 1–9 (2012)
4. Parsons, T.D.: Neuropsychological Assessment Using Virtual Environments: Enhanced Assessment Technology for Improved Ecological Validity. In: Brahmam, S., Jain, L.C. (eds.) *Advanced Computational Intelligence Paradigms in Healthcare* 6. SCI, vol. 337, pp. 271–289. Springer, Heidelberg (2011)
5. Rizzo, A.A., Pair, J., Graap, K., Treskunov, A., Parsons, T.D.: User-Centered Design Driven Development of a VR Therapy Application for Iraq War Combat-Related Post Traumatic Stress Disorder. In: *Proceedings of the 2006 International Conference on Disability, Virtual Reality and Associated Technology*, pp. 113–122 (2006)
6. Parsons, T.D., Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. *Cyberpsychology and Behavior* 11, 17–25 (2008)
7. Nadolne, M.J., Stringer, A.Y.: Ecologic validity in neuropsychological assessment: Prediction of wayfinding. *Journal of the International Neurophysiological Society* 7, 675–682 (2001)
8. Waller, D., Hunt, E., Knapp, D.: The transfer of spatial knowledge in virtual environment training. *Presence* 7, 129–143 (1998)
9. Golledge, R.G.: Cognition of physical and built environments. In: Garling, G., Evans, G.W. (eds.) *Environment, Cognition and Action: An Integrated Approach*, pp. 35–62. Oxford University Press, NY (1991)
10. Thorndyke, P.W., Hayes-Roth, B.: Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology* 14, 560–589 (1982)
11. Hahm, J., Lee, K., Lim, S.L., Kim, S.Y., Kim, H.T., Lee, J.H.: A study of active navigation and object recognition in virtual environments. *Annual Review of CyberTherapy and Telemedicine* 4, 67–72 (2006)

12. Ramloll, R., Mowat, D.: Wayfinding in virtual environments using an interactive spatial cognitive map. In: IV 2001 Proceedings, pp. 574–583. IEEE Press, London (2001)
13. Walker, B.N., Lindsay, J.: The effect of a speech discrimination task on navigation in a virtual environment. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 1536–1541 (2006)
14. Meilinger, T., Knauff, M., Bulthoff, H.H.: Working memory in wayfinding: A dual task experiment in a virtual city. *Cognitive Science* 32, 755–770 (2008)
15. Parsons, T.D., Iyer, A., Cosand, L., Courtney, C., Rizzo, A.A.: Neurocognitive and psychophysiological analysis of human performance within virtual reality environments. *Studies in Health Technology and Informatics* 142, 247–252 (2009)
16. Allanson, J., Fairclough, S.H.: A research agenda for physiological computing. *Interacting with Computers* 16, 857–878 (2004)
17. Parsons, T.D., Cosand, L., Courtney, C., Iyer, A., Rizzo, A.A.: Neurocognitive Workload Assessment using the Virtual Reality Cognitive Performance Assessment Test. In: Harris, D. (ed.) EPCE 2009. LNCS (LNAI), vol. 5639, pp. 243–252. Springer, Heidelberg (2009)
18. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology: Heart rate variability: Standards of measurement, physiological interpretation and clinical use. *Circulation*, 93, 1043–1065 (1996)
19. Ripley, B.D.: *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge (1996)
20. Parsons, T.D., Rizzo, A.A., Buckwalter, J.G.: Backpropagation and regression: Comparative utility for neuropsychologists. *Journal of Clinical and Experimental Neuropsychology* 26, 95–104 (2004)
21. Saltelli, A.: Global sensitivity analysis: An introduction. In: Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output, Santa Fe, New, Mexico, pp. 27–43 (2005)
22. Chen, H., Kocaoglu, D.F.: A sensitivity analysis algorithm for hierarchical decision models. *European Journal of Operational Research* 185, 266–288 (2003)
23. Winebrake, J.J., Creswick, B.P.: The future of hydrogen fueling systems for transportation: An application of perspective-based scenario analysis using the analytic hierarchy process. *Technological Forecasting and Social Change* 70, 359–384 (2003)
24. Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G.: Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors* 40, 79–91 (1998)
25. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors* 45(4), 635–675 (2003)

# A Study of Navigation and Selection Techniques in Virtual Environments Using Microsoft Kinect®

Peter Dam<sup>1</sup>, Priscilla Braz<sup>2</sup>, and Alberto Raposo<sup>1,2</sup>

<sup>1</sup> Tecgraf/PUC-Rio, Rio de Janeiro, Brazil  
peter@tecgraf.puc-rio.br

<sup>2</sup> Dept. of Informatics/PUC-Rio, Rio de Janeiro, Brazil  
{pbraz, abraposo}@inf.puc-rio.br

**Abstract.** This work proposes and studies several navigation and selection techniques in virtual environments using Microsoft Kinect®. This device was chosen because it allows the user to interact with the system without need of hand-held devices or having a device attached to the body. This way we intend to increase the degree of virtual presence and, possibly, reduce the distance between the virtual world and the real world. Through these techniques we strive to allow the user to move and interact with objects in the virtual world in a way similar to how s/he would do so in the real physical world. For this work three navigation and three selection techniques were implemented. A series of tests were undertaken to evaluate aspects such as ease of use, mental effort, time spent to complete tasks, fluidity of navigation, amongst other factors for each proposed technique and the combination of them.

**Keywords:** 3D Interaction, Virtual Reality, Gesture Recognition, HCI.

## 1 Introduction

Virtual Environments, due to enabling realistic and immersive experiences, have seen an increase in importance. Its use in areas such as games, simulation and training, medicine and architectural visualization has pushed the visualization technologies to rapid evolution. However, the way we interact with these environments hasn't evolved as fast, leaving a noticeable gap and hindering the interaction capabilities, since many inherently tri-dimensional tasks have been performed using technologies developed primarily to solve bi-dimensional tasks.

The objective of this work is to propose and study techniques that allow the user to interact in a complete manner using only corporal movements to perform tasks in a virtual environment, especially training and simulation, where the user normally needs to navigate through a scene and interact with equipment. For this, three selection and three navigation techniques have been proposed using Microsoft Kinect® as an input device. These techniques use corporal gestures, most of which aim to keep a certain fidelity to the respective actions in the real world in attempt to increase the naturalness of tri-dimensional interaction.

This paper is organized the following way: section 2 speaks of related work, section 3 presents the proposed techniques, section 4 presents results and analysis of user tests, and section 5 brings the conclusion.

## 2 Related Work

There are several researches in the virtual environment interaction area, but very few of those, up to the current date, make use of Microsoft Kinect®, due to it being a relatively new technology. For this reason, the study of related work focused on work about interaction in virtual environments.

According to Bowman and Hodges [1], interaction in virtual environments is divided into three types: locomotion (navigation), selection and manipulation, where, in many cases, the last two are combined, but can be dissociated. Since in this work both locomotion and selection have been considered, researches about either case have been considered in related work.

**Selection.** Sibert and Jacob [2] present a selection based on gaze direction. It is based upon a directional ray controlled by the direction of the eyes' gaze, eliminating the need of hand-held devices or devices attached to the user. The selection is triggered when the gaze rests upon an object for a certain amount of time. The idea of relating time to selection intention is contemplated in the *Hover* technique, presented in this paper. Rodrigues et al. [3] studied the advantages of applying multi-touch interface concepts in virtual reality environments by mapping 3D space into a virtual touch screen. To enable this, they proposed a wireless glove which is worn by the user and tracked by a specific configuration of Nintendo WiiMote® controllers. The index finger's position is tracked, mapping the axes into system coordinates. The X and Y axes are used to control the mouse cursor on the screen, while the Z axis is used to determine selection intent by establishing a threshold in the real world as if it were a screen. If the finger passes beyond this threshold the selection is activated and a command is triggered, sending haptic feedback, present in the glove. Even though the glove was designed for and tested in 2D interfaces, it inspired the *Push* technique, specifically the gesture of passing an imaginary plane in front of the user to confirm selection (or generating a "click"); and, consequently, also inspired the *Hold* technique.

**Navigation.** One technique that consists in putting the foot in a certain position to navigate is the *Dance Pad Travel Interface*, proposed by Beckhaus, Blom and Haringer [4]. This technique consists of a physical platform (created for the game *Dance-Dance Revolution*), which has directional buttons. The user steps on these buttons and a displacement is created in the direction represented by these buttons. To control the viewing direction the user steps on the directional arrows. One of the navigation techniques proposed in this work (*Virtual Foot Dpad*) was inspired by the *Dance Pad Travel Interface*. During the development of this technique a very similar technique was found in the game *Rise of Nightmares* for the XBOX/Kinect console.

Bouguila, Ishii and Sato [5] created a physical device, similar to a platform, which detects the user's feet and, when moved a certain distance away from the center, activate movement in that direction. To control the viewing direction the user turns his whole body in the desired direction. Because of this, a portion of the user's field of view might not be occupied by the viewing screen, so the device slowly rotates to align the user to the screen again. This work inspired the idea of allowing the user to completely leave a virtual circle, creating a movement vector with origin in the circle's center in the direction of the user's position. This led to the creation of the *Virtual Circle* technique.

### 3 Proposed Techniques

The proposed techniques use information obtained from Microsoft Kinect® as the only data input device. OpenNI [6] was used for communication between the device and the system.

#### 3.1 Selection Techniques

First a virtual hand was developed to follow the user's hand movements in the real world. Moving this virtual hand over objects in the scene enables selection of this object, however the gesture required to select the object depends on which technique is being used. Unlike in Bowman and Hodges [7], due to our work focusing on selection and not exactly manipulation, we did not find the "lever" problem, where the object is attached to the extreme of a selection ray, making it difficult to properly manipulate the object.

**Hover.** This technique is based on the idea that the user will focus her/his attention on an object when s/he wishes to select it [2]. When the user wishes to select an object s/he needs to hover with the virtual hand over that object. A timer will appear and, once emptied, the object will be selected (Fig. 1). When the virtual hand intercepts a selectable object a "pre-counter" is started, introduced to avoid the "Midas Touch" effect, described by Jacob et al. [8]. This allows the user to freely move the virtual hand without actually triggering many visual timers all the time.

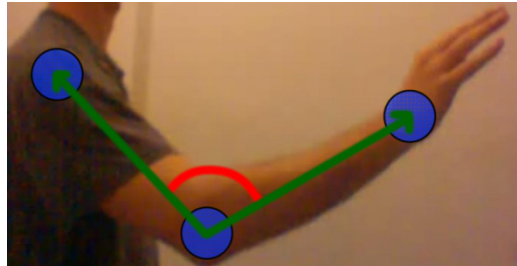
There are two ways to de-select an object with this technique. The first requires the user to move the virtual hand away from the selected object and, after a short time, it will be de-selected. This may not be possible if the object is attached to the virtual hand on all 3 axes, so a second de-selection method was created. The second method requires the user to overlap both hands, which will start a timer to confirm the intention of de-selection and, consequently, de-select the object once the timer runs out.

**Push.** The idea for this technique came from having a virtual plane in front of the user, described by Rodrigues et al. [3]. The user stretches her/his arm and, once it passes a certain threshold, the selection is triggered. The user must then withdraw her/his arm and may interact with the object. To release the object s/he repeats the gesture.



**Fig. 1.** Hover technique timer

The gesture of stretching the arm is detected through the arm's angle, more specifically the angle between the vectors formed by the elbow to the wrist and the elbow to the shoulder, as seen in Fig. 2. Once the angle reaches a pre-established limit, the system activates the selection (or de-selection). One problem present in this technique, described by Rodrigues et al. [3], is the involuntary movement along the X and/or Y axes while the user performs gesture of stretching her/his arm. This problem is more noticeable in cases where interaction requires a higher precision or when the object to be selected is very small on the screen, but for larger objects this problem rarely is an issue.



**Fig. 2.** Arm openness angle

**Hold.** This technique is based on the previous one, as an alternative. Selection is activated in this technique when the user stretches her/his arm, but, unlike the previous one, s/he must maintain the arm stretched during the interaction. De-selection is done by withdrawing the arm.

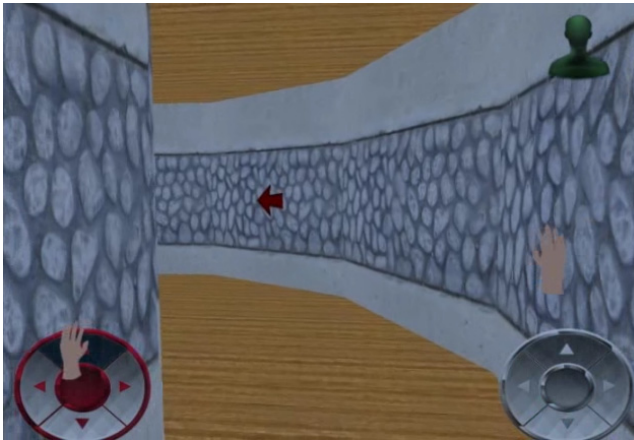
### 3.2 Navigation Methods

For a complete interaction experience the user must be allowed to select and to navigate through the scene. To enable this, three navigation techniques were created. Two of the proposed techniques use *Body Turn* to control the view point orientation. *Body*

*Turn* is a sub-part of these techniques and consists of the user turning her/his shoulders in the direction in which s/he wishes to rotate the view point, while maintaining the central direction of the body facing the screen. This allows the user to control the view and movement direction without the screen exiting her/his field of view.

**Virtual Foot DPad.** This technique was inspired by the work of Beckhaus, Blom and Haringer [4], where they created a physical platform on which the user steps on directional arrows to move in the corresponding direction. The idea was to make a virtual version of this platform. Three joints were used to achieve this: torso, left foot and right foot. The distance of each foot to the torso is calculated and, once one of the feet reaches a certain distance a movement is generated in that direction. This technique uses the previously described *Body Turn* to allow the user to control the view point orientation.

**Dial DPads.** Based on first person games for touch screen devices, such as *iPhone* and *iPad*, this technique uses dials that the user interacts with using virtual hands (Fig. 3). The idea is that it works in a fashion similar to a touch screen, but in larger scale and, instead of using fingers on a screen, the user uses hands. Two dials are displayed on the screen, one in each inferior corner. To the left is the movement control dial and to the right is the view point orientation dial. The user places her/his hand over the dial and stretches the arm to activate it.



**Fig. 3.** Dial DPads controls

**Virtual Circle.** In this technique the system needs to store the position from which the user started the interaction and generates a virtual circle at this spot. The circle is fixed and the user can be compared to an analog joystick. To move in any direction the user simply moves in that direction enough to leave the virtual circle. A vector is then created from the center to the circle to the user's current position, defining direction and speed of the movement (Fig. 4). To stop the movement the user steps back into the circle. For view point orientation the technique uses *Body Turn*.



Fig. 4. Virtual Circle movement vector

## 4 Evaluation and Analysis of Test Results

### 4.1 Evaluation

Selection and navigation tasks were identified for the tests in a 3D virtual environment to exercise the interaction techniques being evaluated.

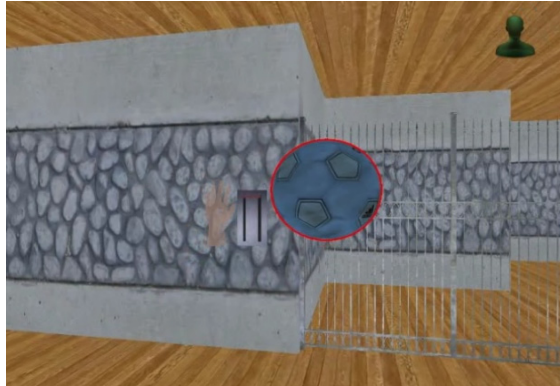
Three use scenarios were defined for execution of the tasks and evaluation of the interaction techniques, described below.

**Scenario 1.** In the first scenario only navigation was contemplated, alternating between the three navigation techniques proposed in this work. This scenario was a corridor, with two  $90^\circ$  curves and a section with a U-turn. The user needed to reach the end of this course, where there would be a red light. Once close enough to this light it would turn off and the user needed to turn around and go back to the initial point.

**Scenario 2.** In this scenario only selection was tested, alternating between the three selection techniques proposed in this work. In this scenario the user had a control panel placed in front of him/her containing a series of levers and buttons (Fig. 1). The user needed to first press several buttons following a specific order, according to which one was lit. After that a series of three red levers needed to be dragged up or down a track to a specific point and released once the indicator showed an acceptable position. At last, two green levers needed to be manipulated simultaneously until the end of their respective tracks.

**Scenario 3.** In this scenario navigation and selection were evaluated, alternating between the navigation and selection techniques. For this test we discarded *Dial Dpads*, because this technique makes use of hands, potentially creating conflict with the three selection techniques. The other two navigation techniques were used in combination with the three selection techniques, creating a total of six combinations. Each of these combinations were tested. This scenario tested the proficiency with buttons and levers, besides a new task: carry a ball while navigating and interacting with other objects at the same time (Fig. 5).





**Fig. 5.** User carrying a ball while navigating in Scenario 3

The order of the tests was changed for each user to avoid that learning had any influence in the general result of the test. In total 9 users were evaluated during the tests using the same physical set up: a room with enough space for free movement with a single large screen.

## 4.2 Analysis of the Results

**Navigation.** Mental effort reflects the degree of interaction fidelity of each technique. *Virtual Circle* had the greatest degree of interaction fidelity and, consequently, demanded less mental effort from the users. Similarly, *Virtual Foot*, which had the second greatest degree of interaction fidelity, demanded greater mental effort.

Comparing one leg of the path amongst the three techniques (Fig. 6) it is possible to observe that the users had a considerably better performance during the U-turn when using *Virtual Circle*. However, to walk in a straight line they performed better with *Virtual Foot*. The reason behind this is that *Virtual Circle* is completely analogical, so if the user moves slightly to any side the movement vector will not be 100% parallel to the walls, creating a slight deviation to one of the sides. This is visible in the initial part (from starting point until the first curve).

**Selection.** The repetition of the gesture for selection and de-selection, present in the *Push* technique, did not please the users, who had trouble with that. *Hover*, on the other hand, was criticized for introducing a delay to be able to select an object, being the least immediate of the three techniques. Despite this, *Hover* was the preferred technique in all tasks. Oppositely, *Push* was the worst in the opinion of the users.

It was made clear that for tasks that require high precision, such as the case of the red levers, the involuntary movement along the X and Y axes highly hinders the interaction, consequently affecting the users' preference of the technique.

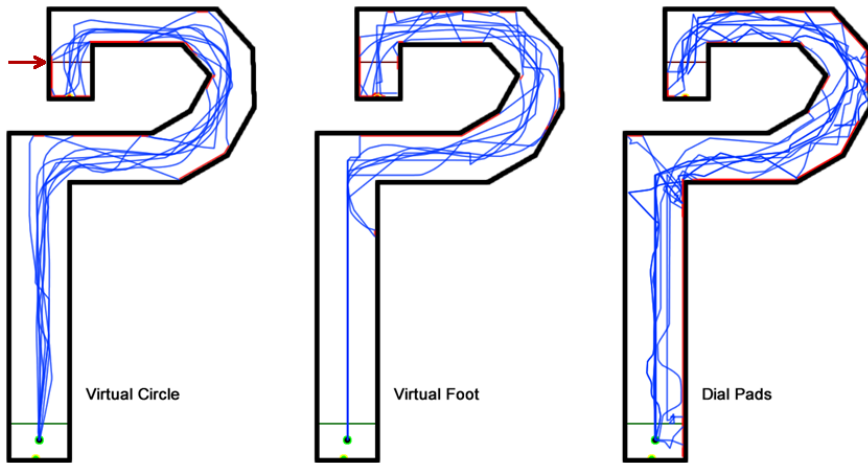


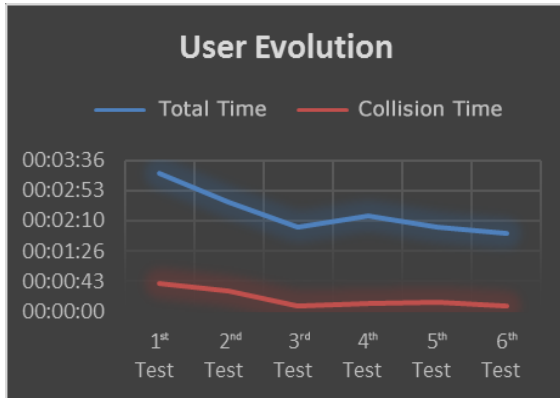
Fig. 6. Path outline for the first leg of the course

Curiously in selection, contrary to navigation, the technique with least interaction fidelity was the one the users preferred. Bowman et al. [6] speak about interaction fidelity, questioning if a technique with higher interaction fidelity means it is necessarily better.

**Combination of Navigation and Selection.** When comparing directly the navigation techniques, we observed that the *Virtual Circle* technique was, in fact, considered slightly better in pair with selection, while the mental effort was very similar, showing that the change in navigation techniques did not have great impact on selection. However, it is possible to observe that strictly comparing navigation tasks, the users preferred *Virtual Circle*.

The technique that had most user technical faults (executing actions by mistake) was *Hold*, with large difference to the second placed technique *Push*. *Hover* did not have any mistakes of this type. These errors were caused by the user withdrawing her/his arm when s/he shouldn't have.

Fig. 7 shows the average execution time for the tasks, considering the order in which they were performed, not sorted by technique. The average time was considered for each 1<sup>st</sup> task of all users, then for each 2<sup>nd</sup> task, and so on. The completion and collision timings show that, no matter which technique combination used, there is a learning curve, indicated by the decreasing lines for task completion. The 4<sup>th</sup> task causes an increase in completion time compared to the 3<sup>rd</sup> task. This is due to changing the navigation technique: the first three tests were applied using one of the navigation techniques, then the last three were applied using a different technique.



**Fig. 7.** User evolution based on average test execution times

The combination of navigation and selection consolidated *Hover* as the overall preferred technique by the users. This preference happened because the users felt more secure to carry objects while navigating, since they didn't have the risk of accidentally dropping the ball.

Besides that, *Virtual Circle* continued to be the preferred navigation technique, but not as evidently as in the first scenario. This was because, in the third scenario, the user was less prone to collision since the environment was more ample than the first, which had narrow corridors.

## 5 Conclusion and Future Work

One of the advantages initially predicted with these techniques was the possibility of interacting with both hands at the same time, a possibility not currently easily supported by current devices. To evaluate this advantage, amongst others, as well as limitations imposed by the techniques, we had to develop user tests. Through these tests we identified which techniques allow a satisfactory interaction, enabling the user to perform tasks in a virtual environment, such as exploring and interacting with objects (despite not being able to rotate and scale them, the users could select and move them).

It was possible to observe that there is clearly a learning curve and, after several tasks, the users would discover ways to use the techniques in which they felt more comfortable. Even though no techniques, in general, had a poor performance, each user, in the end, felt more comfortable with a certain navigation and selection technique. Despite this, it was not possible to compare these techniques with techniques the users were already familiar with, due to the possibility of using both hands simultaneously.

At last, it was possible to verify that Microsoft Kinect® enables the creation of techniques with high degree of interaction fidelity that allow several user actions in a virtual environment in a comfortable manner, besides increasing the user's virtual presence. After some improvements, especially in the implementation of the

techniques, we believe that they can be used in virtual reality applications to control a character and, possibly, to perform more complex tasks than currently possible, mainly due to the possibility of using both hands simultaneously.

**Acknowledgements.** Tecgraf is an institute mainly supported by Petrobras. Alberto Raposo thanks CNPq for the individual grant (process 470009/2011-0).

## References

1. Bowman, D., Hodges, L.: Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages and Computing* 10(1), 37–53 (1999)
2. Sibert, L., Jacob, R.: Evaluation of Eye Gaze Interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2000), pp. 281–288. ACM, New York (2000)
3. Rodrigues, P., Raposo, A., Soares, L.: A Virtual Touch Interaction Device for Immersive Applications. *The Int. J. Virtual Reality* 10(4), 1–10 (2011)
4. Beckhaus, S., Blom, K., Haringer, M.: Intuitive, Hands-free Travel Interfaces for Virtual Environments. In: *New Directions in 3D User Interfaces Workshop of IEEE VR 2005*, pp. 57–60 (2005)
5. Bouguila, L., Ishii, M., Sato, M.: Virtual Locomotion System for Human-Scale Virtual Environments. In: Proceedings of the Working Conference on Advanced Visual Interfaces (AVI 2002), pp. 227–230. ACM, New York (2002)
6. OpenNI, <http://openni.org/>
7. Bowman, D., Hodges, L.: An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D 1997), p. 35. ACM, New York (1997)
8. Jacob, R., Leggett, J., Myers, B., Pausch, R.: An Agenda for Human-Computer Interaction Research: Interaction Styles and Input/Output Devices. *Behaviour & Information Technology* 12(2), 69–79 (1993)

# Legibility of Letters in Reality, 2D and 3D Projection

Elisabeth Dittrich<sup>1</sup>, Stefan Brandenburg<sup>2</sup>, and Boris Beckmann-Dobrev<sup>3</sup>

<sup>1</sup> Research training group Prometei, Berlin, Germany  
elisabeth.dittrich@zmms.tu-berlin.de

<sup>2</sup> University of Technology of Berlin, Chair of cognitive psychology  
and cognitive ergonomics, Berlin, Germany  
stefan.brandenburg@tu-berlin.de

<sup>3</sup> University of Technology of Berlin,  
Chair of Industrial Information Technology, Berlin, Germany  
bdobrev@mailbox.tu-berlin.de

**Abstract.** Virtual prototypes are essential for engineers to understand the complex structures and arrangements of mechatronic products like automobiles. Currently, Virtual Environments (VE) are used for visual analysis and interaction with virtual models. In the next years more supplementary information will be integrated in the VE, completing the 3D-model. This includes names of single parts, corresponding materials or masses. However, up till now there is little explicit research on the psychological effects of additional text visualization in VE's. For example it unclear if it is possible to visualize the textual information like on paper prints or on 2D displays. The current study empirically compares these types of different output mediums to advise rules for visualization of text in 3D Virtual Environments. Results show, that textual information has to be slightly enlarged for the 3D Virtual Environment. In addition, subjects performed better in conditions with projected textual information compared to real text.

**Keywords:** 2D and 3D text, Virtual Environments, legibility of letters, information visualization.

## 1 Introduction

Virtual Reality is a growing technology in different areas of science and industry. In the engineering context it is an “[...] effective tool for a number of purposes.”[1]. Moreover, in the industrial engineering process Virtual Reality is the key technology to validate mechatronic products in early stages of the engineering lifecycle [2]. Its importance becomes evident in virtual design reviews. Within such a review engineers can interact with a three-dimensional model of the desired product in its original size [3]. There the product reacts like a real one. Hence the simulation of products in an early stage of the engineering process helps to find a lot of mistakes in the construction, the dependency of functions and problems with the arrangement of modules. To get there the product engineering lifecycle starts with the definition of product requirements. These requirements are used to design the initial virtual

prototype with geometrical masses and nearly real functions. This virtual prototype can be tested and validated in an immersive virtual environment (VE) [4].

This so called virtual design review is an in-progress validation of the product. It controls the state of the construction and the realization of the requirements [5][6]. Hence this review needs a lot of essential information, for example text-documents, the product-structure, sketches and descriptive statistics [7]. To date each piece of information is displayed on a single resource. Whereas the virtual prototype is mostly realized in an immersive VE, documents, sketches or descriptive statistics are printed or on a separate computer. Therefore participants of a virtual design review constantly need to switch between different sources of information. In addition they cannot interact with the textual information in the design review. For example, viewing and manipulating components of the product-structure inside of the VE would be helpful to access the desired information without interrupting the review. Thus, the visualization of the 3D-model and additional information on separate screens complicates and lengthened the procedure of a virtual design review significantly.

Therefore, research has started to find out how to integrate words in virtual environments (see [8] for an overview).

### **1.1 Information Rich Virtual Environments (IRVE)**

The integration of words in virtual scenes is called Information Rich Virtual Environments (IRVE) [9]. Polys [10] define an IRVE “[...] as a class of visual analysis tools for integrated information spaces [that] start[s] with realistic and perceptual information and enhance it with abstract and temporal information.” [10, p.31]. Therefore a more or less complex data structure is visualized as a three-dimensional object and enhanced with additional information, so called annotations. Annotations might be simple labels, graphs, or other multimedia [11].

As Pick [7] point out, it is much of a challenge to include supplementary information to the virtual object under investigation. Several problems come up when visually integrating annotations to visualized objects. For example, unfavorable annotation positions might occlude data or vice versa [7]. Moreover, if annotations are placed in IRVEs they need to be clearly structured to avoid information clutter or confusion by subjects.

Additionally, [11] provides an overview regarding the parameters that affect the quality of textual annotations in IRVEs. These parameters are: color, fonts, size, background and transparency. Polys and Bowman [11] even conclude that the legibility of words is the most important attribute of annotations in IRVEs. Finally Polys [8] explicitly highlights the size of annotations as a major determinant with respect to their legibility.

### **1.2 Text Legibility**

The international organization of standardization (ISO) defines text legibility as the „[...] ability of unambiguous identification of single characters or symbols that may be presented in a non-contextual format.“ [12, p.25]. To ensure legibility a lot of parameters are to be considered. In ISO [13] ten parameters describing among other

things the contrast, size, thickness of lines of the letters. One of them is the symbol height. The minimum height for digital symbols is determined at 16 minutes of arc, which is larger than the minimum for paper prints. This difference in minimum character size is justified by the fact that the legibility of characters on digital displays is also influenced by “[...] pixel density or resolution, contrast and character font and matrix, as well as viewing distance.” [13, p.12].

However, in an IRVE text legibility is also affected by its resolution. Finally in a virtual environment like a CAVE the viewing distance is dynamic [8], because the position of the viewer is changing while watching the virtual object or text. Therefore this dynamics influences the perception of the virtual objects resolution [11]. Empirical studies showed that reading time of words increased when text was rotated, or the viewing position changed in the IRVE [14], [15]. Additionally this effect is greater for smaller than for larger fonts.

### 1.3 Perception in (IR)VEs

Stereoscopic view in reality means that the two pictures from the two eyes influence the impression of an object. One of the depth cues is the disparity of these two eye pictures. Regarding the effect of disparity on object perception one can state that the wider the disparity, the closer is the focused object. For simulating this three dimensional impression in virtual environments the flat image of the object is projected twice on the screen and the distance both pictures (= disparity) specifies the perceived depth [16].

However, compared to reality the stereoscopic projection in IRVEs might result in the convergence-accommodation-conflict that is connected to simulator sickness [17] and viewing fatigue [18]. In reality people set their focal point of view directly on the object. We obtain a tree-dimensional image because both eyes set this focal viewpoint slightly different. In contrast in a virtual environment, the projected object has no depth itself and people need to focus on the display. However the computer varies the disparity between the two projections of an object in this way that the focus point of the object lays in front of or behind the screen. This difference between the physical (i.e. the screen) and the virtual (i.e. the object) point of focus leads to the convergence-accommodation conflict which in turn is related to an increase in users strain and a decrease in information acquisition.

## 2 Research Question

Previous studies investigated a lot of parameters that affect perception in Virtual Environments. In most of them resolution and the depending font size are not explicitly listed, sometimes not mentioned at all. Moreover there are no standards for text legibility in (IR)VEs. For example, [13] defines larger letter sizes for digital (2D) letters compared to print media. In addition, we pointed out that VEs impose special demands on the reader regarding text legibility.

Hence we want to find out, which minimal resolution and font size is needed to ensure legibility in VEs. Based this theoretical background we would assume that if the

same letters are visualized in normal 2D projection and in a stereoscopic 3D projection, than subjects' performance with respect to legibility is better in the 2D condition. In other words, letter size needs to be larger in 3D compared to 2D. To test this hypothesis we chose a two-step procedure. First, we exactly assessed subjects' eyesight using a real and a projected standard eyesight examination plate. Second, we tested whether resolution needs to be different for 2D and 3D projection of letters.

### **3 Empirical User-Study**

#### **3.1 Sample**

A total number of  $N = 21$  subjects participated in this experiment, 13 (61%) females and 8 (39%) male. Subjects age ranged between 14 and 60  $M = 31.5$  ( $SD = 13$ ) years. Most of them (71%) were students, 1 person (5%) was a retiree and another 5 participants (24%) were employees. Almost half of the sample (47%) had corrected to normal vision. All of them wore glasses and no contact lenses. Their mean correction was  $M = 1.53$  ( $SD = 2.03$ ) diopter. Eight of these 10 participants with corrected to normal vision showed near-sightedness. One participant had color deficiency, one was night-blind and another one had dyslexia.

#### **3.2 Technical Equipment**

In the experiment we used a portable powerwall. This powerwall was operated through an active stereo projector (DepthQ HDs3D) with 1280\*720 ppi at 120 Hz. The stimulus material was rear-projected on a silver screen that was optimized for 3D environments. The projector setup was also optimized for higher pixel density. This means it was possible to realize 1 mm pixel height and thus a picture of 1280\*720 millimeters. Two computers with Nvidia Quadro FX 3800 graphic cards presented the stimulus material. The visualization software for the stereoscopic view was VDP by ICIDO (now ESI-group). The shutter glasses and the shutter emitter were named APG600. The eye chart was designed in NX (Siemens PLM) and modulated with Deep Exploration.

#### **3.3 Experimental Design**

The first independent variable was the kind of letter projection, with two manifestations: a) normal 2D and b) stereoscopic 3D projection. All settings were done in the software for 3D visualization. For the 2D condition the disparity was adjusted on zero percent. Therefore the stimulus material and all other conditions were absolutely the same in both conditions.

The second independent variable was the number of pixels per letter (resolution). In both conditions, 2D and 3D, the resolution of letters differed in 6 steps. For each step of this manipulation the virtual distance of the plate was decreased for one meter. The algorithm for the 3D illusion of a rearward moved letter made them smaller and their disparity higher, so the letters seemed to be more behind the screen. In total we



moved the letters in six steps rearward, each one resembled a virtual meter. Each time we placed the letter one more virtual meter backwards; the participant went one physical meter forward. Therefore the letter size was constant (see also minute of arc in Table 1) and the amount of pixels for one letter decreased. The following table shows the six steps of reducing the resolution from the best to the worst resolution.

The dependent variable was text legibility, which is dichotomous with the manifestations non legible and legible. Hence a border of legibility for every participant and condition was quantified. This border served as the dependent variable for subsequent analysis.

**Table 1.** Manipulation of the number of pixel (second column), distance between the participant and the projection screen (3<sup>rd</sup> column) and (in last column) the post hoc calculated minute of arc of the letters

	number of pixels	distance to screen	post hoc calculated minute of arc
1. best resolution	11	6 m	6,3°
2.	9	5 m	6,1°
3.	7	4 m	6,0°
4.	6	3 m	6,8°
5.	5	2 m	8,5°
6. worst resolution	4	1 m	13,7°

Contrary to our expectations table 1 shows that the virtual and the physical meter were not equivalent, because the depending minutes of arc are not constant. We would consider this as a bug of the virtual visualization software. It remains unclear why the algorithm produced this error.

### 3.4 Procedure

First of all participants completed the questionnaire assessing demographic variables. Now, subjects completed the pre-test. The goal of this pre-test was to check the sightedness of the participants. Hence all subjects completed a standardized eye examination procedure using a plate (see Fig.1 on the right side). For this eye examination, the standardized procedure is to read each line of letters from top to the bottom of the plate. The line of letters where subjects were not able to read the characters anymore defined the sightedness in percent.

After finishing reading the letters of the physical eye examination plate, participants repeated the same eye examination procedure with a projected plate of letters (see middle of Fig.1). Here they were placed 20 centimeters closer to the projection screen to compensate the different distances of both plates. Every participant accomplished both tasks in the same order and the eyesight in percent was noted for both conditions. However this second trial of the pre-test was to guarantee transferability from the physical to the projected examination plate. Therefore no differences in eyesight were expected between both conditions (pre-test hypotheses).

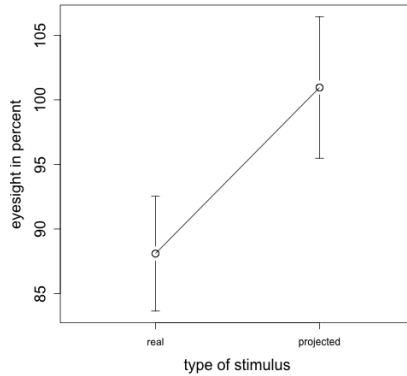


**Fig. 1.** Shows the physical standardized eye examination plate on the right hand side and the projected plate on the left hand side

In the experimental part of this study we asked participants again to read out loud one line of letters of the 2D projected examination plate. More in detail, they were requested to read the line of letters that resembled 100% eyesight. Each participant repeated this procedure (i.e. reading the line of letters that resembled 100% eyesight) 16 times, 8 times in the 2D condition and 8 times in the 3D condition. We designed different digital plates, so that letters in the same line were randomized from the whole alphabet. The participants began with the best resolution in the 2D condition. When the subjects spelled the presented letter correctly, the next smaller resolution was shown. In case of an error, the measurement stopped and the experimenter noted the number of pixels. This procedure was repeated eight times. Trials starting from a decreasing number of pixels were alternately presented to trials increasing beginning with an increasing number of pixels. After finishing all 8 trials for the 2D condition, we added disparity and repeated the procedure for the 3D condition.

## 4 Results

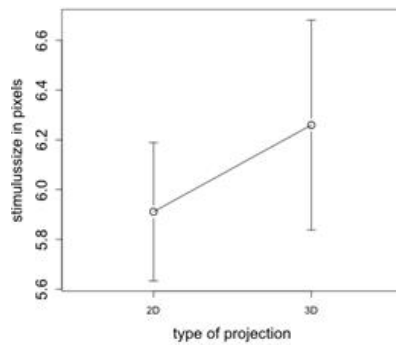
A paired t-test was computed for the two hypotheses, the difference between the real and the projected eye-examination plate and the difference between 2D and 3D stereoscopic view. In addition confidence intervals (CI) and effect sizes ( $d$ ) after [19] were reported if applicable. Regarding the pre-test question, we assumed that there is no difference in subjects' performance depending on the real and the projected eye chart. In contrast to the hypothesis the statistical test did reveal a medium and significant difference between the real and the projected version of the chart,  $t(1,20) = -3.23$ ,  $p = 0.004$ ,  $d = 0.7$ . As visualized in figure 2, subjects eyesight increased about  $M = 12.85\%$  (95% CI ranging from 4.57% to 21.14%) when changing from a real eye chart to the projected version.



**Fig. 2.** Subjects' eyesight depending on the type of eye chart used. Bars indicate the 80% confidence intervals.

With respect to the research question, we obtained a tendency for the difference between the 2D and the 3D visualization of letters,  $t(1,20) = 1.57$ ,  $p = 0.06$ ,  $d = 0.2$ . Figure 3 shows that stimulus size increased about  $M=0.4$  pixels (6.7%).

For practical reasons we would suggest to report the recognition border in whole numbers. Therefore the 2D-legibility-border is  $M = 5.9$  pixels, which resembles 6 pixels height for one letter. The 3D-legibility-border is  $M = 6.3$  pixels which should be rounded to 7 pixel height of a small letter.



**Fig. 3.** Subjects' recognition border depending on the type of projection. Bars indicate the 80% confidence intervals.

## 5 Discussion

Text in Virtual Environments is needed in engineering contexts. For example in virtual design reviews information like measurements of a distance or properties of the designed objects are discussed. If at all, a small fraction of this textual information is mostly arranged near by the object in form of an annotation. However up till now it

was not clear how large these annotations need to be. Hence we were interested in the general question how the legibility of the text depends on the resolution and which font size in VEs ensures text legibility. Results show that subjects' performance is better when letters are a little bit larger in stereoscopic projection compared to 2D projection. However, statistics revealed a small tendency and not a very strong effect. Possibly this result might depend on the rather small amount of participants that were used. In addition, it is possible that the difference between 2D and 3D conditions has other reasons. For example, in the 3D condition participants have worn shutter-glasses. This might have altered their performance in the letter-reading task. Moreover we might have elicited a ceiling effect in terms of the readability of letters at all. Hence the difference between 2D and 3D projection would have been larger under different circumstances. Results of related studies undermine this position. For example [20] discovered that a font size of 9 pixels is optimal for the legibility of virtual lowercase letters. In our study, it was around 6 pixels and therefore a lot lower than Sheedys' numbers [20]. Possibly other parameters of text legibility like luminance, contrast or thickness of lines of the letters were very well set in our experiment. Therefore subjects' showed best performance and the difference of type of projection was at its minimum. Besides related work the result of our pre-test also supports this position. In contrast to our hypothesis, subjects' eyesight was better in the 2D projection compared to the real eye-examination plate. On the one hand this result points out the potential for projections to improve peoples' sightedness. On the other hand, we might have shown the importance of comparable testing situations. We invested a great amount of effort regarding the standardization of subjects' eye examination. Of course we tried to balance light and contrast conditions for the use of the real and projected eye examination plate. Again, results indicate that lighting conditions, contrast or other indicators of legibility was better for the projected eye examination compared to the real one. The direction of the difference between real world and projected test-plate contradicts the definition of larger fonts for digital texts compared to print media. However, 2D and 3D projection was extremely comparable since we used the same method and instruments for the visualization of letters for both projections.

Finally practical implications and limitations of our work need to be mentioned. Regarding the practical implications we were lucky to find a small difference between 2D and 3D projection only. Since the amount of information that is used in a virtual design review is usually quite large [7], small fonts enable engineers to include much information in the Virtual Environment. However, this advantage directly relates to the problems of text visualization in Virtual Environments. As stated above, unfavorable annotation positions might occlude data or vice versa [7]. Moreover, since we found that annotations do not need to be large in terms of their size, programmers might be tempted to include many of them in IRVEs. This would directly lead to information clutter or confusion of subjects. Additionally, other parameters of annotation in IRVEs might reduce generalizability of our results. For example if programmers chose unfavorable colors, fonts, backgrounds or transparency settings a small increase in font size might not be enough to ensure text legibility. In line with [15] and [11], we would assume these text legibility-determining factors to interact with

each other. With respect to these interactions many open questions are left at the end of this work. Future research should therefore replicate our small effect of the 2D versus 3D comparison with a larger sample. If it is possible to replicate our finding, one could think of expanding the experimental setting to more realistic examples. In the present work, we only used extremely simple stimuli with maximum contrast, black letters on white background. Contrasting this approach, real world mockups are colorful. And so are annotations. On a more complex level of research one could further address questions of occlusion or visual clutter [8].

To sum up, a lot of rules for legibility are accessible for printed documents. However, there are none for stereoscopic texts. In contrast to current practice we would propose to only slightly enlarge (plus one pixel) text size in information rich virtual environments. Hence practitioners would be enabled to include larger amounts of data in their virtual environments. Hence, the present work is important since it shows three things. First, texts only need to be a little larger in (IRVEs) than in 2D. Second, projected text bears the potential to enhance subjects' performance compared to real world stimulus material. And third, we did not obtain any evidence for the convergence-accommodation-conflict [17], [18].

## References

1. Wickens, C., Holland, J.: *Engineering Psychology and Human Performance*. Prentice Hall, New Jersey (2000)
2. Krause, F.-L., Franke, H.-J., Gausemeier, J.: *Innovationspotenziale der Produktentwicklung*. Carl Hanser Verlag, München (2007)
3. Schmid, D.: *Konstruktionslehre Maschinenbau*. Verlag Europa- Lehrmittel, Haan-Gruiten (2009)
4. Blümel, E., Straßburger, S., Sturek, R., Kimura, I.: Pragmatic Approach to Apply Virtual Realty Technology in Accelerating a Product Life Cycle. In: *Proceedings of the International Conference INNOVATIONS*, pp. 199–207 (2004)
5. Kamiske, G.F., Ehrhart, K.J., Jacobi, H.-J., Pfeifer, T., Ritter, A., Zink, K.J.: *Bausteine des innovativen Qualitätsmanagement*. Hanser, München (1997)
6. Masing, W.: *Handbuch Qualitätsmanagement*. Hanser, München (1994)
7. Pick, S., Hentschel, B., Wolter, M., Tedjo-Palczynski, I., Kuhlen, T.: Automated positioning of annotations in immersive virtual environments. In: Kuhlen, T. (ed.) *Joint Virtual Reality Conference of EGVE-EuroVR-VEC*, pp. 1–8. The Eurographics Association (2010)
8. Polys, N.: *Display Techniques in Information-Rich Virtual Environments*. Doctoral dissertation, Virginia Polytechnic Institute and State University (2006)
9. Bowman, D., Hodges, L., Bolter, J.: The Virtual Venue: User-Computer Interaction in Information-Rich Virtual Environments. *Presence: Teleoperators and Virtual Environments* 7(5), 478–493 (1998)
10. Polys, N., Bowman, D., North, C.: The role of Depth and Gestalt cues in information-rich virtual environments. *International Journal of Human-Computer Studies* 69, 30–51 (2011)
11. Polys, N., Bowman, D.: Design and display of enhancing information in desktop information-rich environments: challenges and techniques. *Virtual Reality* 8, 41–54 (2004)
12. ISO 9241-302: *Ergonomics of human-system interaction – Terminology for electronic visual displays*. ISO, Geneva (2008)

13. ISO 9241-303: Ergonomics of human-system interaction – Requirements for electronic visual displays. ISO, Geneva (2011)
14. Larson, K., van Dantzich, M., Czerwinski, M., Robertson, G.: Text in 3D: some legibility results. In: CHI 2000 Extended Abstracts on Human Factors in Computing Systems, pp. 145–146. ACM, New York (2000)
15. Grossman, T., Wigdor, D., Balakrishnan, R.: Exploring and reducing the effects of orientation on text readability in volumetric displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 483–492. ACM, New York (2007)
16. Goldstein, B.E.: Wahrnehmungspsychologie. Spektrum, Heidelberg (2001)
17. Häkkinen, J., Pölönen, M., Takatalo, J., Nyman, G.: Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations. In: Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services, pp. 227–230. ACM, New York (2006)
18. Stern, A., Javidi, B.: Three dimensional image sensing, visualization and processing using integral imaging. Proceedings of the IEEE 94(3), 591–607 (2006)
19. Cohen, J.: Statistical power analysis for the behavioral sciences, 2nd edn. Erlbaum, New York (1988)
20. Sheedy, J.E., Subbaram, M.V., Zimmerman, A.B., Hayes, J.R.: Text legibility and the letter superiority effect. Human Factors: The Journal of the Human Factors and Ergonomics Society 47(4), 797–815 (2005)

# The Visual, the Auditory and the Haptic – A User Study on Combining Modalities in Virtual Worlds

Julia Fröhlich and Ipke Wachsmuth

AI & VR Lab, Faculty of Technology, Bielefeld University  
Universitätsstraße 25, 33615 Bielefeld, Germany  
{jfroehli, ipke}@techfak.uni-bielefeld.d

**Abstract.** In order to make a step further towards understanding the impact of multi-modal stimuli in Virtual Reality we conducted a user study with 80 participants performing tasks in a virtual pit environment. Participants were divided into four groups, each presented a different combination of multi-sensory stimuli. Those included real-time 3D graphics, audio stimuli (ambient, static and event sounds), and haptics consisting of wind and tactile feedback when touching objects. A presence questionnaire was used to evaluate subjectively reported presence on the one hand, and on the other physiological sensors were used to measure heart rate and skin conductance as an objective measure. Results strongly indicate that an increase of modalities does not automatically result in an increase of presence.

**Keywords:** Presence, User Study, Multi-modal Feedback, Virtual Reality.

## 1 Introduction

Ever since Morton Heilig developed the Sensorama Simulator [8], multi-sensory feedback has been claimed to be of notable importance. Half a century later the quality of graphical presentation has increased dramatically, but many virtual reality applications fall short on presenting multi-sensory experiences. Even worse, the stimuli present are sometimes in conflict with the virtual world (e.g. a silent virtual room but the air condition in the Lab is noisy).

One of the major goals in Virtual Reality is to create a highly immersive environment. Modern hardware facilitates real-time 3D graphics. In 3D-setups like the CAVE [3], the user is located directly in the virtual world and becomes part of it. However, there are other factors which influence the user's immersion, for instance natural interaction and navigation. Moreover the feeling of being in the world is also influenced by the way the user experiences the world with other senses. In order to make a further step towards understanding the correlation between multi-sensory stimuli and the perceived presence we conducted a user study.

In the following we start by giving an overview of related work. Our focus will be on research of immersion and presence tied to Virtual Reality applications and multi-sensory stimuli and their impact. The second part of the paper will describe the conducted user study, starting with an overview of the setup and procedure, followed by results. A discussion and some directions for future work will sum up our contribution.

## 2 Related Work

Presence has been defined as “a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences” [18]. An immersive virtual world helps users to accomplish their tasks in an efficient way: it facilitates building a mental model of the environment [13]. Moreover, existing mental models of interaction in the real world can quickly be adapted to those needed in the virtual environment. There are different opinions about how to maximize immersion. Sheridan suggested three essential factors as follows [15]:

1. The quality (and quantity) of visual, auditive and haptic feedback
2. The possibility of moving the point-of-view and the sensors in a virtual environment
3. The possibility of changing the environment, as easy as in the real world

The second factor can be regarded as accomplished. By combining a CAVE [3] and a tracking system, the user is able to move around freely (in the limited space of the CAVE) and the viewpoint adapts to his position in real-time. However, the first and third factor are only partly accomplished. Visual quality of immersive worlds is at a point of nearly being photo-realistic. Other modalities like auditive and haptic feedback are continuously enhanced, but still, in many cases there is a lack of multi-modality.

Acoustic and tactile feedback are used more commonly and there are many different concepts and devices. The integration of wind is not as common. An additional wind setup was e.g. implemented by Deligiannidis and Jacob [4]. It was used to improve speed perception as the user was navigating in a 3D-world with a scooter. Since they had a fixed wind direction their setup was limited to specific scenarios. But yet, they conducted a user study which proved not only a higher reported presence but an objectively better task performance.

Measuring presence is still a challenging task and many different types of experimental setups have been proposed. The most common measure is reported presence through questionnaires. Over the years several questionnaires have been developed, in particular the Witmer-Singer [18] and Slater-Usuh-Steed [17] questionnaires have been used in numerous studies. Those questionnaires are considered *subjective* measures, because different persons may respond totally different to the same environment. Therefore researchers are looking for more *objective* measures as well. Physiological reactions (heart rate, skin conductance, and skin temperature) were tested in virtual environments in order to find a correlation



with reported presence [10]. For certain stress-inducing environments this correlation was significant. From the results of the visual cliff studies by Gibson and Walk [7] in 1960 the idea of a virtual pit evolved. The initial experiments showed that the presence of a cliff is a fear-evoking experience and for most people it requires a huge amount of willpower to cross a precipice like this. First introduced by Mel Slater and colleagues in 1995 [16] – virtual pits are a commonly used test scenario in virtual worlds today. They evoke a physiological reaction and therefore facilitate the availability of an objective measure of presence.

To investigate which influence further modalities have on the perceived presence, Dinh et al. [5] conducted a user study in 1999. It indicated that an increased amount of modalities results in increased perceived presence and memory of objects in the environment. The environment was presented with different combinations of multi-sensory stimuli consisting of head-mounted graphic display, auditory, tactile, and olfactory cues. Each modality had two levels of realism. Results showed significant effects for auditory and for tactile cues. For olfactory cues a non significant trend was measured. Surprisingly the quality of the visual cues had no impact on the perceived presence. The authors argued that additional sensory cues, except for visual ones, work in a simple additive fashion on the sense of presence. Whether this still holds for today's virtual reality applications, is a subject of the study presented in the sections to follow.

### 3 Experimental Setup

In this section we describe the experimental setup used for the study. First we present the hard- and software setup followed by a discussion of the used navigation and interaction method. Furthermore a description of the virtual world as presented in the study is given.

#### 3.1 Setup

Our setup consists of a 3-sided CAVE-like environment. The user wears tracked glasses for dynamical adaption of the viewpoint. Furthermore a sound and a wind setup are employed to generate multi-sensory stimuli.

**Spatial sound** is realized with eight speakers (one in each corner) and two subwoofers underneath the floor. Sounds are divided into three different types. *Ambient sounds* represent a base level of output which is more or less constant over a larger region of a scene. As long as the user is within the defined area, ambient sound will be played without direction and always at the same volume. In addition, this concept allows the definition of environmental properties which influence the audio rendering, to fit the environment, such as an outside scenario, a cave or a concert hall. *Static sounds* are directly coupled to an object. They are adjusted in volume and direction with regard to their position relative to the user. *Event sounds* are only triggered when the related event occurs (e.g. a ball hitting the floor).

**Tactile feedback** is accomplished by ART fingertracking devices which track the thumb, the index finger and the middle finger. At the tip of each finger

three wires made of memory metal are attached which shorten when heated momentarily. When repeated in short intervals a vibration is created, which can be utilized as haptic feedback to the user. The strength of feedback can be regulated steplessly.

**Wind effects** are accomplished by eight controllable fans, which are located at the upper bound of the projection area. In consideration of available space, costs, as well as fine-grained adaptation of wind direction we chose a setup in which the fans are mounted evenly distributed on a nearly circular arrangement. The fans were chosen with a special focus on being as silent as possible.

### 3.2 Navigation and Interaction

There are many different interaction methods in Virtual Reality, but most of them aim for efficiency instead of realism. The possibility of fast manipulation of objects is often more important than intuitiveness and ease of use. Still, a user will naturally try to grasp an object directly. After manipulating, the user will expect the object to fall down on the ground, like in the real world. That is why a natural hand interaction method was chosen in our scenario.

In most CAVE setups realistic walking is not feasible. Thus, it has to be replaced by a less intuitive navigation metaphor, but still aiming for increased immersion. A study by Slater et al. [16] indicated that walking-in-place resulted in a higher subjective sense of presence than a push-button-fly (along the floor plane) navigation. Instead of tracking the users head movements to indicate whether they are walking or not, we thus track the feet directly with markers. For rotation, we make use of users' head orientation. Sudden viewpoint changes are interpreted as changes of walking direction. Similarly to the original paper [16], the viewpoint is then dynamically adjusted until the user faces the front wall again [14].

### 3.3 Environment

The experiment subsequently described took place in a virtual pit environment. The presented world as seen in Figure 1 was employed in the study. It consisted of two rooms, a training room (right) and the virtual pit room (left). The virtual world was designed based on *Instantreality*, a consistent, platform-independent framework for fast and efficient application development for Virtual and Augmented Reality [6].

Our test scenario is designed similar to other virtual pit environments (e.g. [11]) with some adjustments. Since we used a CAVE and not a head-mounted display (HMD) as most similar studies did, the training room was constructed bigger. This is to give users a chance to try the walking-in-place navigation. The training room was furnished as a living room and offered enough details to spend time on exploring the environment. Furthermore we supplied some objects (a ball and a gong) to train natural hand interaction. The pit room was not furnished but contained only the pit with a small gallery and two planks. The pit was actually



**Fig. 1.** The virtual environment as presented in the user study

covered with a virtual glass floor, therefore it was possible to walk right over it (actually nearly nobody did this).

The presentable modalities were:

- step sounds when walking (steps on a wooden floor)
- a radio playing the theme from 'the Good, the Bad and the Ugly'
- a pretty loud event sound when hitting a gong
- drop sounds when a ball fell on the ground
- a mechanical sound when the door to the pit room was opened
- atmospheric wind sound when entering the pit room
- tactile feedback when touching objects
- haptic wind blowing from the open windows into the pit room.

## 4 User Study

To measure the effect of the presented modalities on users' presence we conducted a study with 80 participants. Participants were recruited through postings in the university building and they were rewarded with chocolate. The only mandatory requirements were that participants had no significant fear of heights, did not participate in any of our previous studies, and were native speakers of German.

Participants were divided into four groups each presented a varied combination of modalities. As a baseline all groups were presented the same graphical world: a virtual pit setup with a training room and a pit room as described in sect. 4.3. The first group did not get any further modalities besides the graphical one, while the second group had additional acoustic feedback whereas the third group had additional haptic feedback. The fourth group was presented the full combination of visual, auditory, and haptic stimuli.

## 4.1 Questionnaires

There were six types of questionnaires used in this study as described in the following:

1. A questionnaire asking for demographic information.
2. The Immersion Tendency Questionnaire (ITQ) by Witmer and Singer [18]. It consists of 12 questions to measure the capability of individuals to get immersed in daily activities like reading or watching a movie.
3. The Simulator Sickness Questionnaire [9] – given before and after the study – to measure influence of the virtual trip on participants' health condition.
4. The two height anxiety questionnaires as introduced by Cohen [2] consisting of 20 situations which evoke a fear of heights.
5. A presence questionnaire similar to the University College London (UCL) Presence Questionnaire, also known as the Slater-Usoh-Steed (SUS) Questionnaire [17]. It consists of 13 questions concerning the overall experiment.
6. A questionnaire with open questions about participants' experience. The questions were designed to check if the presented modalities were noticed and if they were appropriate. Moreover participants had the chance to write a few lines about how they liked the experiment and if there were any improvements suggested.

## 4.2 Procedure

Before entering the training room participants had to answer five questionnaires: demographic information, Immersion Tendency Questionnaire, Simulator Sickness Questionnaire and the two height anxiety questionnaires. Then a calibration of markers for gesture recognition was performed to ensure the same conditions for each participant.

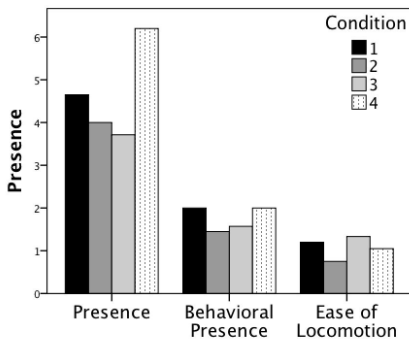
All participants started with a training procedure of about 15 minutes. Step-by-step, they learned to look around, to walk and to interact with objects in the virtual world. The training included a walk through the whole training room, hitting the gong and throwing a ball. Afterwards participants were asked to proceed to the pit room and throw two balls at a target on the ground of the pit. After the part within the virtual world was completed, another set of questionnaires was given to the participants: the Simulator Sickness Questionnaire as before, the UCL Presence Questionnaire, and the questionnaire with open questions about their experiences including memory questions. In addition we recorded physiological data through heart rate and skin conductance sensors, in order to measure the physical reaction to the virtual pit objectively. The whole procedure took about 60 minutes for each participant.

## 5 Results

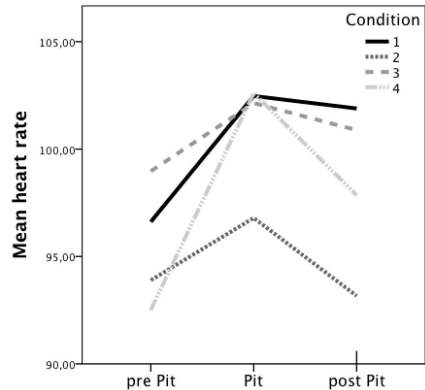
As described above there were two measures of presence: subjective and objective. First we will present the results of the subjective measure from the UCL

questionnaire, which consists of 13 questions. Participants were to answer the questions on a 7-point Likert scale. The topics covered are: the sense of being in the Virtual Environment (VE), the extent to which the VE becomes the dominant reality, and the extent to which the VE is remembered as a place. Three of those 13 questions are used to measure the reported *behavioral presence* based on studies indicating that behaviors as a response to stressful stimuli and reported behavioral presence correlate. Therefore a participant would react with more pit-avoidance, the more real the precipice would seem. Another three questions measure reported *ease of locomotion* – the ability to navigate effortlessly in the virtual world. The remaining seven questions measure the reported *presence*: the "sense of being" in a place or environment (e.g. a virtual environment) even when one is physically situated in another [18].

Usuh et. al [17] suggested to count the number of high answers (top 3 on a 7 point Likert scale) for the corresponding questions. Therefore presence is rated between 0 and 7 (number of possible high answers), whereas behavioral presence and ease of locomotion range from 0 to 3. Figure 2 shows the mean for each of the measures divided into the four groups. An analysis of variance for the UCL questionnaire showed highly significant results for the presence measure (Welch’s  $F(3, 40.39) = 8.893, p < .001$ ).



**Fig. 2.** Evaluation of the mean reported presence, behavioral presence and ease of locomotion. **Condition 1:** Graphics, **Condition 2:** Graphics & Audio, **Condition 3:** Graphics & Haptics, **Condition 4:** Graphics & Audio & Haptics



**Fig. 3.** Mean heart rate before, during, and after exposure to the pit room (measured in bpm). **Conditions 1, 2, 3,** and **4** see left.

Bonferroni’s post-hoc comparisons of the four conditions showed that participants in the second condition (Graphics & Sound) ( $M = 4.00, SD = 2.03, CI[3.05, 4.95]$ ) as well as participants in the third condition (Graphics & Haptics) ( $M = 3.71, SD = 1.31, CI[3.12, 4.31]$ ) rated the presence significantly lower than participants in Condition 4 (all modalities) ( $M = 6.20, SD = 0.89,$

**Table 1.** Overview of the results from the presence related measures

	Condition 1	Condition 2	Condition 3	Condition 4
Presented modalities	visual	visual auditory	visual haptic	visual auditory haptic
Presence	4.65 (2.16)	4.00** (2.02)	3.71** (1.31)	6.2** (1.91)
Behavioral Presence	2,00 (0.86)	1.45 (0.89)	1.57 (0.87)	2.00 (1.03)
Ease of Locomotion	1.2 (1.15)	0.75 (0.91)	1.33 (1.11)	1.05 (1.15)
Heart rate increase	5.85 (9.63)	2.90* (4.22)	3.17* (4.66)	10.10* (8.57)
Skin conductance increase	3.22 (2.51)	2.52 (1.66)	1.73 (1.04)	3.21 (3.11)

$CI[5.78, 6.62]$ ),  $p < .001$ . Condition 1 ( $M = 4.65$ ,  $SD = 2.16$ ,  $CI[3.64, 5.66]$ ) was rated lower as well and is significant for  $p < .05$ . Behavioral Presence showed a similar non significant trend. Measures for Ease of Locomotion showed no significant differences across the four conditions.

As for the subjective measures of presence, the same analysis was performed for the measured heart rate and skin conductance. Due to equipment malfunction the physiologic data of six participants are missing. Figure 3 gives an overview of the measured heart rates for each group in each phase. 'Pre pit' is the whole training phase, 'pit' is the time from entering until leaving the pit room, and 'post pit' is measured for three minutes from leaving the pit room until the end of the experiment. In order to compare heart rates for training (pre pit) and pit room we calculated the increase for each group. The mean during the training phase and the mean while in the pit room were calculated and compared. An ANOVA showed significant results ( $F(3, 70) = 4.1$ ,  $p < .05$ ). Bonferoni post-hoc comparisons of the four groups indicate that heart rate increase in the fourth condition ( $M = 10.1$ ,  $SD = 8.57$ ,  $CI[5.97, 14.22]$ ) was significantly higher than participants' heart rate increase in Condition 2 ( $M = 2.89$ ,  $SD = 4.22$ ,  $CI[0.87, 4.93]$ ),  $p < .05$  and 3 ( $M = 3.17$ ,  $SD = 4.66$ ,  $CI[0.85, 5.48]$ ),  $p < .05$ . No significant difference but a similar trend was measured for skin conductance.

Table 1 gives an overview of these results. Mean and standard deviation (in parentheses) are given. For heart rate and skin conductance the increase from the training to the pit room is given. Significant values are marked with asterisks ( $* = p < .05$ ;  $** = p < .01$ ).

In addition correlation analyses were conducted for the presence related measures. The Score on the Immersion Tendency Questionnaire correlates with the reported presence ( $r = .24$ ,  $p < 0.05$ ) and behavioral presence ( $r = -.29$ ,  $p < 0.05$ ). Furthermore the score for reported *behavioral* presence correlates with gender (negative for male ( $r = -.32$ ,  $p < 0.01$ )). The time spent playing

computer games did correlate with gender (for male  $r = .25, p < 0.05$ ) but not with any results from the reported or observed presence related measures.

## 6 Discussion and Conclusion

Our results suggest that more presented modalities do not necessarily result in an increased perceived presence. In this study participants tended to rate presence lower when only presented with one additional (audio or haptic) cue. The recorded physiological data support this observation. When presented the full combination of modalities – the visual, the auditory, and the haptic – the perceived presence is significantly higher.

While this at first may seem counter-intuitive, it might be an uncanny valley effect in today's virtual reality applications. As Masahiro Mori first stated for the robotics domain, if human replicas look and act almost, but not perfectly, like actual human beings, it causes a response of revulsion among human observers [12]. He called it a valley, corresponding to the valley on the graph of the comfort level of humans as a function of a robot's human likeness. Brenton et al. described the same valley in the domain of virtual characters. If such characters are too close to a human but not perfect, people tend to dislike them [1]. A similar dip can be seen in our presence results (fig. 2), where Conditions 2 and 3 represent the valley.

One explanation could be that with all stimuli combined users' expectations are better met and an increase in reported presence is observed. When presented only one additional cue, users may have expected more. Due to the overall availability of technology today, people are used to have a high amount of presented stimuli. For example, state-of-the-art computer games present at least very good auditory cues, and a lot of them tactile feedback as well. It may be not enough to present only a few sounds to make the world's overall believability better. Our results suggests that the enhancement of virtual worlds with multi-modal stimuli, does not work in a simple additive fashion (or not anymore), like concluded by Dinh et al. [5].

Thus our future work will focus on how to improve users' presence with multi-sensory stimuli. Doing so, further improvement of multi-modal feedback will be an important aspect, since it should be a relevant factor for the enhancement of immersion. There will be an effort to overcome the uncanny valley to increase the overall believability of and therefore make a step towards significantly improved immersion with additional multi-sensory stimuli. One approach could be the enrichment of virtual worlds with even more stimuli, for example heat or smell. Further the influence of "autonomy" factors, like the availability of an intelligent virtual agent as an interaction partner, seems worthwhile investigating as well.

## References

1. Brenton, H., Gillies, M., Ballin, D., Chatting, D.: The uncanny valley: does it exist. In: 19th British HCI Group Annual Conference: Workshop on Human-Animated Character Interaction (2005)

2. Cohen, D.C.: Comparison of self-report and overt-behavioral procedures for assessing acrophobia. *Behavior Therapy* 8(1), 17–23 (1977)
3. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., Hart, J.C.: The cave: audio visual experience automatic virtual environment. *Commun. ACM* 35(6), 64–72 (1992)
4. Deligiannidis, L., Jacob, R.J.K.: The vr scooter: Wind and tactile feedback improve user performance. In: *3DUI 2006*, pp. 143–150 (2006)
5. Dinh, H.Q., Walker, N., Song, C., Kobayashi, A., Hodges, L.F.: Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In: *Proceedings of the IEEE Virtual Reality, VR 1999*, pp. 222–228. IEEE Computer Society, Washington, DC (1999)
6. Fellner, D., Behr, J., Bockholt, U.: Instantreality - a framework for industrial augmented and virtual reality applications. In: *The 2nd Sino-German Workshop on Virtual Reality & Augmented Reality in Industry* (2009)
7. Gibson, E., Walk, R.: The visual cliff. *Scientific American* 202, 64–71 (1960)
8. Heilig, M.L.: Sensorama simulator, u.s.patent no.3050870 (August 1962)
9. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilenthal, M.G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology* 3(3), 203–220 (1993)
10. Meehan, M., Insko, B., Whitton, M., Brooks Jr., F.P.: Physiological measures of presence in stressful virtual environments. In: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2002*, pp. 645–652. ACM, New York (2002)
11. Meehan, M., Razaque, S., Insko, B., Whitton, M., Brooks, J., Frederick, P.: Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology and Biofeedback* 30, 239–258 (2005)
12. Mori, M., MacDorman, K., Kageki, N.: The uncanny valley (from the field). *IEEE Robotics Automation Magazine* 19(2), 98–100 (2012)
13. Pausch, R., Proffitt, D., Williams, G.: Quantifying immersion in virtual reality. In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997*, pp. 13–18. ACM Press/Addison-Wesley Publishing Co, New York (1997)
14. Renner, P., Dankert, T., Schneider, D., Mattar, N., Pfeiffer, T.: Navigating and selecting in the virtual supermarket: Review and update of classic interaction techniques. In: *Virtuelle und Erweiterte Realität: 7. Workshop der GI-Fachgruppe VR/AR*, pp. 71–82. Shaker Verlag GmbH (2010)
15. Sheridan, T.: Further musings on the psychophysics of presence. In: *1994 IEEE International Conference on Systems, Man, and Cybernetics, Humans, Information and Technology*, vol. 2, pp. 1073–1077 (October 1994)
16. Slater, M., Usoh, M., Steed, A.: Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.* 2(3), 201–219 (1995)
17. Usoh, M., Catena, E., Arman, S., Slater, M.: Using presence questionnaires in reality. *Presence: Teleoper. Virtual Environ.* 9(5), 497–503 (2000)
18. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. *Presence* 7(3), 225–240 (1998)



# Spatial Augmented Reality on Person: Exploring the Most Personal Medium

Adrian S. Johnson and Yu Sun

Robot Perception and Action Lab (RPAL), University of South Florida, Tampa, FL, USA  
asjohns4@mail.usf.edu, yusun@cse.usf.edu

**Abstract.** Spatial Augmented Reality (SAR) allows users to collaborate without need for see-through screens or head-mounted displays. We explore natural on-person interfaces using SAR. Spatial Augmented Reality on Person (SARP) leverages self-based psychological effects such as Self-Referential Encoding (SRE) and ownership by intertwining augmented body interactions with the self. Applications based on SARP could provide powerful tools in education, health awareness, and medical visualization. The goal of this paper is to explore benefits and limitations of generating ownership and SRE using the SARP technique. We implement a hardware platform which provides a Spatial Augmented Game Environment to allow SARP experimentation. We test a STEM educational game entitled ‘Augmented Anatomy’ designed for our proposed platform with experts and a student population in US and China. Results indicate that learning of anatomy on-self does appear correlated with increased interest in STEM and is rated more engaging, effective and fun than textbook-only teaching of anatomical structures.

**Keywords:** spatial augmented reality, self-referential encoding, education.

## 1 Introduction

In Spatial Augmented Reality (SAR) [1-10], projectors render graphical information onto real objects. SAR has rarely been applied to the body of humans. Recently, tracking human pose in real-time without markers has become readily available using the Microsoft Kinect [11]. Thus, dynamic projection on-person is now widely accessible, motivating study of such interactions.

Further, traditional SAR research has yet to explore specific psychological effects of using the human body as a display medium. Relating digital content and avatars with the self has been found generating a sense of ownership. Ownership is the sense of self-incorporation a user adopts for an avatar, object, or interaction they identify with personally [12, 13, 14, 17].

Once ownership is established, a psychological effect termed Self-Referential Encoding (SRE) holds that information relating to the self is preferentially encoded and organized above other types of information [15, 16]. The Spatial Augmented Reality on Person (SARP) system and application we present here differ from typical SAR research in

that they are designed to explore the effects of self-referential interactions. Our SARP system projects interactions onto and around a user's dynamic moving body.

A hyper-personalized avatar virtualizing the user's very own internal makeup en-joins her every move in spatial as well as temporal unity. The above referenced psychological research has laid groundwork in identifying and quantizing ownership and SRE effects. However, little work has been done to render and study such promising effects in a commonly measured context providing for more formal assessment. The Augmented Anatomy SARP application we developed for our system explores teaching anatomical structures on person to study these effects in learning a practically applicable and standardized core curriculum.

## **2 SARP Platform Description**

A hardware platform that provides a spatially augmented game environment (SAGE) capable of rendering the SARP technique was developed (Figure 6). The hardware consists of ubiquitous technology such as a Microsoft Kinect for tracking and pose estimation, a high-lumens BenQ MX commodity projector for ambient-light resistant visualizations, and a standard computing device with a graphic processing unit to efficiently and intelligently drive the two former I/O components. Standard projector and Kinect calibration techniques provide for spatial unity of avatar and user in the physical world.

Up to two users may freely move and turn within the bounds of the SAGE while interacting with content rendered on their bodies. Anatomical structures remain in sync with the body, maintaining the user's sense of ownership.

## **3 Augmented Anatomy SARP Application Description**

An educational game and learning analysis tool that renders the SARP technique entitled Augmented Anatomy was developed. We provided a gesture-based approach for the user to learn anatomical structures and trigger interactions. Users can turn around and see the anatomy from different angles. The student may select an anatomical structure with their hand to learn more.

During the quiz interaction, the names of anatomical structures are spoken by the computer. The user is given time to touch the correct organ. In two player competition mode, the first player to correctly identify the correct structure wins points. The correct structure highlights before moving onto the next organ to provide feedback.

Several anatomical learning modes are provided. For instance, if the player makes a muscle the muscular system triggers (Figure 3). Crossing the arms into a skull and crossbones position under the head will display the skeletal system (Figure 1). Within each anatomical system mode, the players are automatically quizzed to identify the anatomical structures visible in that system. The system verbally asks the players to "point to" structures by name or functionality based on questions modeled from standardized assessments widely utilized by schools.



**Fig. 1.** Two players in the “Augmented Anatomy” Spatial Augmented Game Environment



**Fig. 2.** Augmented Anatomy SAGE in a classroom setting



**Fig. 3.** The muscular system is triggered via double bicep pose

## 4 Results

We present statistics in the form  $M \pm S$ , where  $M$  is a mean and  $S$  is a standard deviation or  $ES$ ,  $p$  where  $ES$  is an effect size and  $p$  is a probability for significance testing.

### 4.1 Expert Survey

Augmented Anatomy was demonstrated to k-12 teachers ( $n=7$ ) to assess SARP teaching effectiveness potential in a classroom setting. Each later anonymously answered survey questions. All teachers either agreed or strongly agreed the system would be effective ( $M=6.71 \pm 0.45$ ) in class and more engaging ( $M=6.43 \pm 0.49$ ) than a textbook (Figure 4).

Descriptive responses included “capturing attention, high levels of on-task behavior,” and an “increased level of interest.” One teacher touted “fun and immediate feedback’ as a cause and another mentioned ‘the fact that the organs show up on the kids’ shirts makes the activity personal.” One expert responded “it looks fairly easy to use, and is student-centered allowing the student to take part in their learning versus listening to a teacher lecture on the topic.”

### 4.2 Student Survey

User evaluation results ( $n=16$ ) support system was more engaging ( $M=6.0 \pm 1.21$ ) effective ( $M=5.60 \pm 1.50$ ) and fun ( $M=6.36 \pm 0.69$ ) than learning from a textbook.

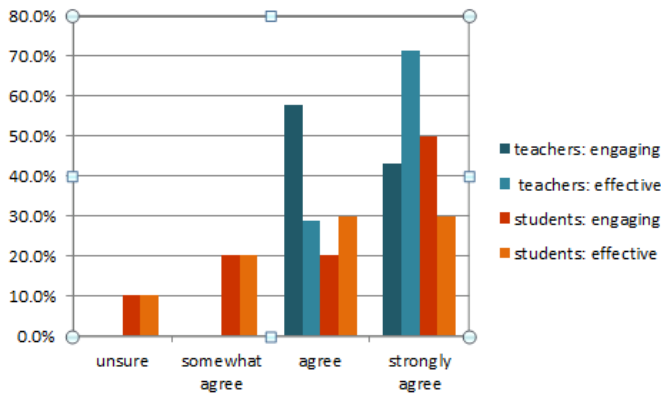
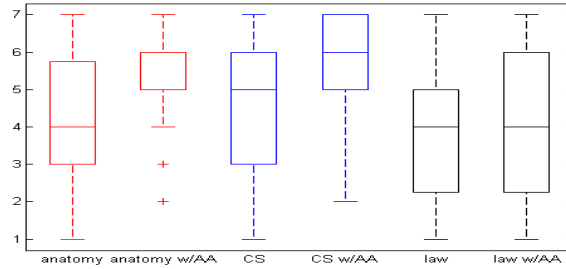


Fig. 4. 100% of teachers, 90% of students agree: more engaging and effective than textbook

### 4.3 Subject Interest Levels

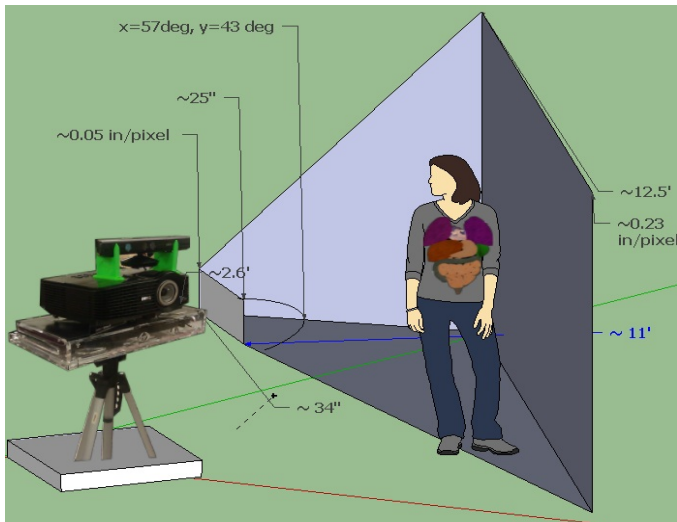
Two groups of students participated in an online STEM subject interest survey. The first group had not seen Augmented Anatomy and the second completed the survey after seeing the system ( $n=81$ ). The system had an immediate positive impact on student’s interest in related STEM subjects. Results found support for an increased interest in anatomy ( $ES=+1.01$ ,  $p<.0048$ ) and computer science ( $ES=+1.16$ ,  $p<.0027$ ) with no significant change for law (Figure 5).



**Fig. 5.** STEM Interest levels significantly increased, law did not

#### 4.4 Learning Effectiveness

We tested effectiveness on (n=50) university and high school students. The subject's education levels varied from 10<sup>th</sup> grade to Ph.D candidate and were evenly distributed among male and female. Incorrect identifications rapidly reduced during successive iterations. 100% identification occurred on average within 2.1 iterations.



**Fig. 6.** Our interactive learning setup and practical ranges using Kinect, projector and laptop

## 5 Concluding Remarks

Here we proposed a new system and application for exploring specific psychological effects of on-person interactions. Our studies have shown positive usability, positive expert feedback, and increase in student engagement. We are conducting tests in real classrooms to determine if SRE, as leveraged by SARP, can increase benchmark scores on standardized material.

**Acknowledgements.** This material is based upon work supported by the National Science Foundation under Grant No. 1035594.

## References

1. Raskar, R., Welch, G., Fuchs, H.: Spatially Augmented Reality. In: First International Workshop on Augmented Reality (September 1998)
2. Bandyopadhyay, D., Raskar, R., Fuchs, H.: Dynamic Shader Lamps: Painting on Real Objects. In: The Second IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), New York, NY, October 29-30 (2001)
3. Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., Forlines, C.: iLamps: geometrically aware and self-configuring projectors. *ACM Transactions on Graphics (TOG)* 22(3) (July 2003)
4. Bimber, O.: *Spatial Augmented Reality: Merging Real and Virtual Worlds*. AK Peters (2005)
5. Azuma, R.: A Survey of Augmented Reality Presence: Teleoperators and Virtual Environments, pp. 355–385 (August 1997)
6. Drascic, D., Grodski, J.J., Milgram, P., Ruffo, K., Wong, P., Zhai, S.: ARGOS: a display system for augmenting reality. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Amsterdam, The Netherlands, April 24-29, p. 521 (1993)
7. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. *Multimed Tools Appl.* 51, 341–377 (2011)
8. Feiner, S.: Augmented reality: a long way off? AR Week, <http://Pocket-lint.com> (archived from the original on March 6, 2011) (retrieved March 3, 2011)
9. <http://www.t-immersion.com/trylive/trylive%E2%84%A2-augmented-reality-fashion>
10. Hagbi, N., Bergig, O., El-Sana, J., et al.: Shape Recognition and Pose Estimation for Mobile Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 17(10), 1369–1379 (2011)
11. <http://www.microsoft.com/en-us/kinectforwindows/>
12. Botvinick, M., Cohen, J.: Rubber hands’ feel’touch that eyes see. *Nature* 391(6669), 756–756 (1998)
13. González-Franco, M., Pérez-Marcos, D., Spanlang, B., et al.: The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In: *VR Conference*, pp. 111–114 (2010)
14. Odom, W., Zimmerman, J., Forlizzi, J.: Teenagers and their virtual possessions: design opportunities and issues. In: *CHI*, pp. 1491–1500 (2011)
15. Symons, C.S., Johnson, B.T.: The self-reference effect in memory: A meta-analysis. *Psychological Bulletin* 121(3), 371 (1997)
16. Baddeley, A.: Working memory. *Science* 255(5044), 556 (1992)
17. Cunningham, S.J., Turk, D.J., Macdonaldet, L.M., et al.: Yours or mine? Ownership and memory. *Consciousness and Cognition* 17(1), 312–318 (2008)

# Parameter Comparison of Assessing Visual Fatigue Induced by Stereoscopic Video Services

Kimiko Kawashima<sup>1</sup>, Jun Okamoto<sup>1</sup>, Kazuo Ishikawa<sup>2</sup>, and Kazuno Negishi<sup>3</sup>

<sup>1</sup> NTT Network Technology Laboratories  
3-9-11, Midori-Cho, Musashino-Shi, Tokyo, 180-8585, Japan

<sup>2</sup> Tokyo Polytechnic University

<sup>3</sup> Keio University

{kawashima.kimiko, okamoto.jun}@lab.ntt.co.jp

**Abstract.** A number of three-dimensional (3D) video services have already been rolled out over IPTV. In 3D video services, there are concerns that visual fatigue still exists, so evaluation of visual fatigue induced by video compression and delivery factors is necessary to guarantee the safety of 3D video services. To develop an assessment method of visual fatigue, we conducted evaluation experiments designed for 3D videos in which the quality of the left and right frames differ due to encoding. We explain the results from our evaluation experiments of visual fatigue, that is, results of specific parameters of visual fatigue biomedical assessment methods.

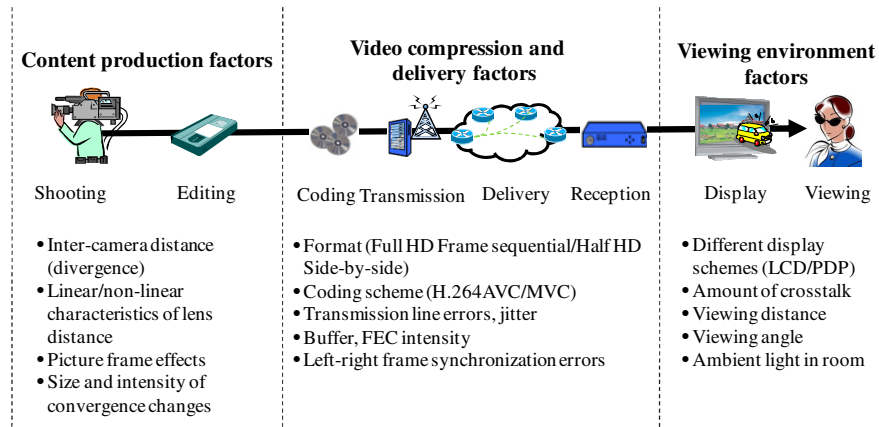
**Keywords:** 3D video, quality assessment, visual fatigue, encoding.

## 1 Introduction

Three-dimensional (3D) movies have become popular worldwide. A number of manufacturers have already put 3D televisions into the market, and the market for 3D-video-related products is surging. 3D broadcasting has been made available in markets around the world, and a number of 3D video services have already been rolled out over IPTV ([1], [2]). 3D broadcasting and 3D video services over IPTV use limited transmission bandwidth (i.e. bit-rate). Therefore, to provide high-quality 3D video services, it is important to compress and transmit information effectively.

Image safety (e.g. visual fatigue and visually induced motion sickness) as well as image quality concerns with conventional 2D video services must be considered for quality of 3D video services. Although 3D movies have become popular, concerns for image safety still exist. For example, the National Consumer Affairs Center of Japan received inquiries and complaints from people regarding visual fatigue after watching 3D movies [3]. Therefore, visual fatigue problems are important issues to consider. For service providers to provide high-quality 3D video services, 3D service design and management based on evaluation of image safety as well as image quality are important.

Visual fatigue is caused by a number of factors such as content production, video compression and delivery, and viewing environment (Fig. 1). Because a number of 3D video services have already been rolled out over IPTV, evaluation of visual



**Fig. 1.** 3D video processing chain

fatigue induced by video compression and delivery factors is necessary to guarantee the safety of 3D video services. The 3D Consortium (3DC) Safety Guidelines for Dissemination of Human-friendly 3D [4] defined safety guidelines for content production and viewing environment factors. Therefore, users are able to watch safe 3D content in a safe viewing environment. However, even if safe 3D content is delivered, video compression and delivery factors may bring about visual fatigue. Video compression and delivery factors are critical for safe 3D video services. A previous study on visual fatigue focused on content production and delivery factors [5, 6, 7]; however, there have been few studies that have focused on video compression and delivery factors. Recently, 3D video service providers have discussed new 3D video compression and delivery methods to achieve higher image quality with lower bit-rate. Service providers are concerned that the difference between the left and right frame quality induces visual fatigue. Therefore, our aim is to evaluate visual fatigue induced by video compression and delivery factors. Using our results of visual fatigue evaluation, we plan to develop video compression and delivery methods to achieve higher image quality and lessen visual fatigue. In addition, as a telecom company, we aim to provide safe 3D video services.

This paper is organized as follows. Previous studies related to evaluation of visual fatigue is explained in Section 2. We discuss the experimental methods and results in Section 3. Finally, the conclusion and further studies are described in Section 4.

## 2 Related Work

There are quality assessment methods for visual fatigue. For example, the Simulator Sickness Questionnaire (SSQ) is used to evaluate visually induced motion sickness and viewing fatigue, and the Visual Analogue Scale (VAS) is used to evaluate relief from fatigue by requiring the consumption of certain healthy food. However, these methods are targeted for people who feel extreme exhaustion. On the other hand, biomedical assessment methods evaluate viewer sensitivity or comfort by measuring their vision. For example, experiments have been conducted on visual fatigue induced



**Table 1.** Biomedical assessment parameters

Parameters
Critical Fusion Frequencies (CFF)
Vision binocular vision, simultaneous perception, position of eye, and fusion
Eye-blink
Pupil constriction rate

**Table 2.** 3D videos

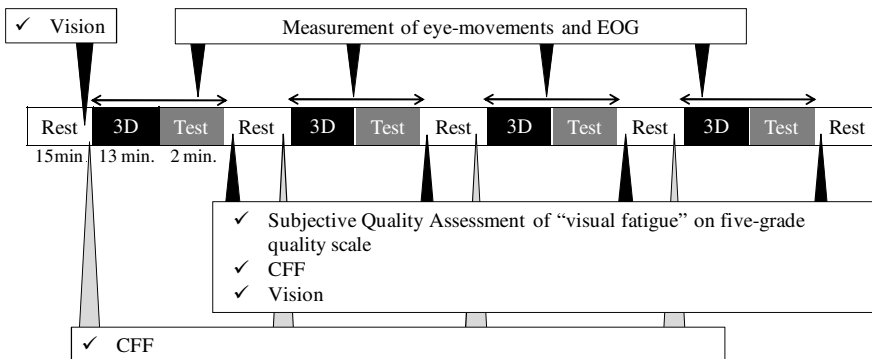
	video 1	video 2	video 3	video 4
Left image quality	Reference	Reference	Reference	Reference
Right image quality	Reference	6 Mbps	3 Mbps	1 Mbps

by 3D monitors and content, and Visual Display Terminal (VDT) tasks using biomedical assessment methods [8-12]. However, no generalized index of visual fatigue has been established. Therefore, our aim was to choose specific parameters of biomedical assessment methods that can be used to assess visual fatigue with a high degree of accuracy, similar to conventional subjective quality assessment methods.

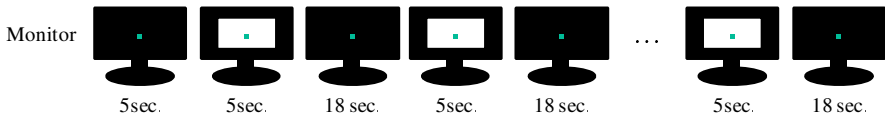
### 3 Visual Fatigue Evaluation Experiments

#### 3.1 Methods

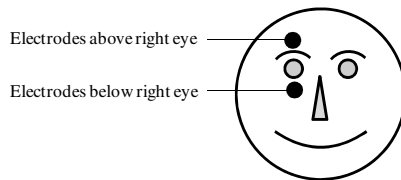
As explained above, we conducted visual fatigue evaluation experiments by using the biomedical assessment methods with several parameters listed in Table 1. These parameters are said to be able to evaluate visual fatigue [4, 9-12]. Our aim was to choose specific parameters of biomedical assessment methods that can evaluate visual fatigue induced by 3D videos in which the quality of the left and right frames differ due to encoding.



**Fig. 2.** Flow of our experiment



**Fig. 3.** Sequence of test video



**Fig. 4.** Schematic representation of electrode locations for EOG recording

In our experiment, 13 participants watched 3D videos of around 13 minutes. Participants viewed each video sequence at a distance of  $3H$  ( $3H$  is about 200 cm,  $H$  indicates the ratio of viewing distance to picture height) from a 46-inch 3D monitor. The participants viewed the 3D video with polarized glasses. The room luminance was 20 lux. We used the 3D content, “2028:Belief” provided by the Digital Content Association of Japan (DCAJ) [13]. This 3D content is produced in compliance with 3DC Safety Guidelines [4] in order to exclude visual fatigue caused by 3D content. This content is about 13-minute drama. We used this content and prepared four kinds of 3D videos based on the hypothesis that the greater the difference between the left and right frame quality in the 3D videos, the stronger the feelings of visual fatigue; one 3D video had the same quality on the left and right frames, and the other three had different quality between the left and right frames due to encoding (Table 2). In Table 2, “Reference” means the source video, and “ $x$  Mbps ( $x=1, 3, 6$ )” means the bit rate of encoding the source video. After watching each 3D video, participants watched the test video in order to measure pupil diameter stable. We show the flow of our experiment in Fig. 2 and the sequence of the test video in Fig. 3. In Fig. 3, the green point was the point of gaze, and the black and white image was presented alternately about 5 times.

We measured eye movement for recording pupil diameters and sight direction, and electro-oculogram (EOG) for counting the number of eye-blinks while participants were watching the 3D videos. We attached electrodes above and below the right eye to record the EOG shown in Fig. 4. We then measured critical fusion frequencies (CFFs) before and after they watched the 3D videos, and vision (e.g. binocular vision, simultaneous perception, position of eye, and fusion) after they finished watching the 3D videos. In addition, participants evaluated their feelings of “visual fatigue” on a five-grade quality scale after they had finished watching the 3D videos. We presented the four 3D videos randomly for each participant.

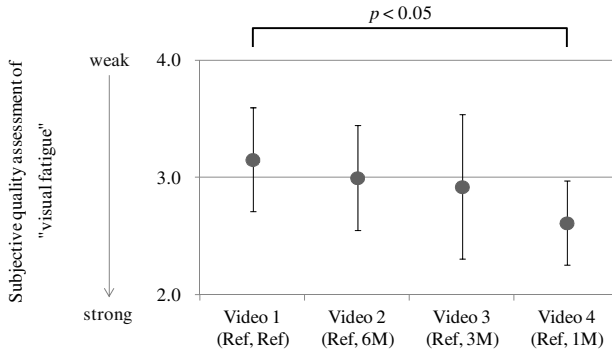


Fig. 5. Results of subjective quality assessment methods of "visual fatigue"

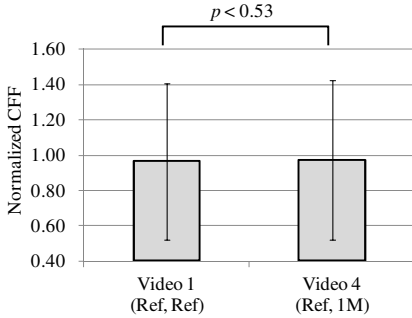
### 3.2 Results

To choose specific parameters of biomedical assessment methods that can evaluate visual fatigue induced by 3D videos in which the quality of the left and right frames differ due to encoding, we compared the biomedical assessment methods' results. In particular, we compared the results of when participants felt strong visual fatigue with those of when they felt weak visual fatigue. We hypothesized that participants felt stronger visual fatigue after they watched video 4 in which there was the largest difference in left and right frames, as described in Section 3.1. In fact, our results of subjective quality assessment of "visual fatigue" showed that the greater the difference between the left and right frame quality in the 3D videos, the lower the grade of subjective quality assessment of "visual fatigue", which means the stronger their feelings of visual fatigue (Fig. 5). We used 13 participants' data. Error bars represented 95% confidence intervals. There was a significant difference of 5% between videos 1 and 4 in the paired t-test.

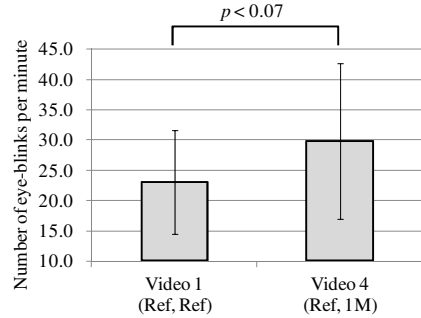
Therefore, we compared the biomedical assessment methods' results when participants watched videos 1 and 4. We then chose the biomedical assessment parameters that could evaluate the visual fatigue of videos 1 and 4 by about 5% significant difference using the paired t-test. This means that the biomedical assessment parameters we chose can match the accuracy of the conventional subjective quality assessment methods of "visual fatigue" described above.

Table 3. Definition of mean CFF values

Measurement timing of CFF	Definition
Mean CFF values before participants watched video $x$ ( $x=1,2,3,4$ )	$CFF_{video\ x\ pre}$ ( $x=1,2,3,4$ )
Mean CFF values after participants watched video $x$ ( $x=1,2,3,4$ )	$CFF_{video\ x\ post}$ ( $x=1,2,3,4$ )



**Fig. 6.** Results of normalized CFF



**Fig. 7.** Results of eye-blinks

### 3.2.1 CFF

In this section, we explain the results of CFF and consider whether CFF can be used to evaluate visual fatigue.

We used 13 participants' data in which there were no loss. We measured the frequencies in which participants could not detect flicker when the frequency was increased from 20 Hz and that in which the participants could detect flicker when the frequency was decreased from 60 Hz frequency. We measured both frequencies three times and calculated the mean CFF values of these frequencies. To reject individual differences, we defined the mean CFF values, which are listed in Table 3. Then, we defined the normalized CFF ( $nCFF_x$ ) in Eq. (1).

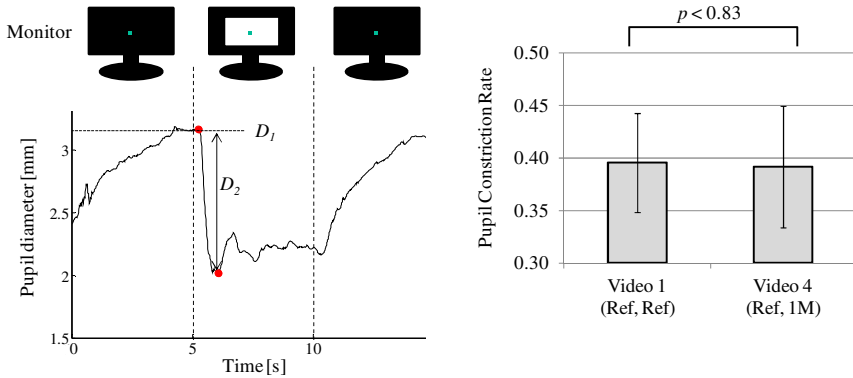
$$nCFF_x = CFF_{video\ x\text{-}post} / CFF_{video\ x\text{-}pre} \quad (x = 1,2,3,4) \quad (1)$$

We show the results of normalized CFFs in Fig. 6. Error bars represented 95% confidence intervals. According to Fig. 6, there was a significant difference of 53% between videos 1 and 4 in the paired t-test. This result did not match the accuracy of the conventional subjective quality assessment methods of "visual fatigue". Therefore, CFF was not able to evaluate visual fatigue induced by 3D videos in which the quality of left and right frames differ due to encoding.

### 3.2.2 Eye-Blink

In this section, we explain the results of eye-blink and consider whether eye-blink can be used to evaluate visual fatigue.

We used 11 participants' data in which there were no loss. Using the results of the EOG measurement, we calculated the number of eye-blinks per minute while they watched 13-minute 3D videos. Our results showed that the greater the difference between the left and right frame quality in the 3D videos, the greater the number of eye-blinks (Fig. 7). Error bars represented 95% confidence intervals. In addition, there was a significant difference of 7% between videos 1 and 4 in the paired t-test. This result matched the accuracy of conventional subjective quality assessment methods of "visual fatigue". Therefore, we determined that eye-blink could be used to evaluate visual fatigue induced by 3D videos in which the quality of left and right frames differ due to encoding.



**Fig. 8.** Analytical terms of pupil measurement **Fig. 9.** Results of pupil constriction rate

In a previous study, the results showed that the number of eye-blinks was higher when watching 3D video than in 2D video [10]. Another study gives the conclusion that the number of eye blinks was higher when viewing the 3D video with moderate visual fatigue than with low visual fatigue [9]. Therefore, our results are consistent with the results of previous studies and eye-blinking is considered as an indicator for measuring visual fatigue.

### 3.2.3 Pupil Constriction Rate

In this section, we will proceed with the analysis of eye-movements and consider whether pupil diameter and other parameters of eye-movements can be used to evaluate the visual fatigue or not.

We used 8 participants' data in which there were no loss. We analyzed in terms of the pupil diameter ( $D_1$ ) and the amplitude of pupillary constriction ( $D_2$ ), as shown in Fig. 8 . Then, we defined the pupil constriction rate ( $CR$ ) in Eq. (2).

$$CR = D_2 / D_1 \tag{2}$$

**Table 4.** Results of parameter comparison

Parameters	Judgments	$p$ -value
CFF	No	$p = 0.53$
Vision binocular vision, simultaneous perception, position of eye, and fusion	No	These did not changed according to the 3D videos that participants watched.
Eye-blink	Yes	$p = 0.07$
Pupil constriction rate	No	$p = 0.83$

We show the results of pupil constriction rate in Fig. 9. Error bars represented 95% confidence intervals. According to Fig. 9, there was a significant difference of 83% between videos 1 and 4 in the paired t-test. This result did not match the accuracy of conventional subjective quality assessment methods of "visual fatigue". Therefore, we determined that pupil constriction rate could not be used to evaluate visual fatigue induced by 3D videos in which the quality of left and right frames differ due to encoding.

### 3.2.4 Results of Parameter Selection for Developing Visual Fatigue Assessment Methods

In this section, we summarize the results of our parameter selection given in Sections 3.2.1, 3.2.2 and 3.2.3. Table 4 listed the results of parameter comparison. According to Table 4, we determined that out of the parameters we measured and analyzed eye-blink is a possible parameter for assessing visual fatigue induced by 3D videos in which the quality of left and right frames differ due to encoding.

## 4 Conclusion

We conducted visual fatigue evaluation experiments by using the biomedical assessment parameters (CFF, vision, eye-blink, and pupil constriction rate) to evaluate visual fatigue induced by video compression and delivery factors. We compared the biomedical assessment parameters' results when participants watched video 1, in which there was no difference in left and right frames, and video 4 in which there was the largest difference in left and right frames. We chose biomedical assessment parameters that can be used to evaluate the visual fatigue with about a 5% significant difference using the paired t-test, and that match the accuracy of the conventional subjective quality assessment methods of "visual fatigue". Out of these parameters, eye-blink is a possible parameter for assessing visual fatigue induced by 3D videos in which the quality of left and right frames differ due to encoding. In the future, we will develop objective visual fatigue evaluation methods based on eye-blink in order to develop video compression and delivery methods to achieve higher image quality with lower bit-rate and lessen visual fatigue.

## References

1. ESPN 3D, <http://espn.go.com/espn/3d/>
2. BBC News - Olympic Games coverage: HD, robotic cameras and 3D, <http://www.bbc.co.uk/news/technology-18690822>
3. NCAC NEWS From National Consumer Affairs Center of Japan 22(4) (November 2010), [http://www.kokusen.go.jp/e-hello/data/ncac\\_news22\\_4.pdf](http://www.kokusen.go.jp/e-hello/data/ncac_news22_4.pdf)
4. 3D Consortium (3DC) Safety Guidelines for Dissemination of Human-friendly 3D, [http://www.3dc.gr.jp/english/scmt\\_wg\\_rep/index.html](http://www.3dc.gr.jp/english/scmt_wg_rep/index.html)

5. Nojiri, Y., Yamanoue, H., Hanazato, A.: Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing. In: Proceedings of SPIE, vol. 5291, pp. 303–313 (2004)
6. Choi, J., Shin, H., Sohn, K.: Smart stereo camera system based on visual fatigue factors. In: 2012 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–5. IEEE Press, Seoul (2012)
7. Emoto, M., Niida, T., Okano, F.: Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *Journal of Display Technology* 1(2), 328–340 (2005)
8. Ukai, K.: Human Factors for Stereoscopic Images. In: 2006 IEEE International Conference on Multimedia and Expo, pp. 1697–1700. IEEE Press, Toronto (2006)
9. Kim, D., Choi, S., Park, S., Sohn, K.: Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks. In: 2011 17th International Conference on Digital Signal Processing, pp. 1–6. IEEE Press, Corfu (2011)
10. Lee, E., Heo, H., Park, K.: The comparative measurements of eyestrain caused by 2d and 3d displays. *IEEE Transactions on Consumer Electronics* 56(3), 1677–1683 (2010)
11. Murata, K., Araki, S.: Accumulation of VDT Work-Related Assessed by Visual Evoked Potential, Near Point Distance and Critical Flicker Fusion. *Industrial Health* 34, 61–69 (1996)
12. Uetake, A., Murata, A., Otsuka, M., Takasawa, Y.: Evaluation of visual fatigue during VDT tasks. In: 2000 IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1277–1282. IEEE Press, Nashville (2000)
13. 2028:Belief (in Japanese),  
<http://www.dcaj.org/news/3D-belief/belief.html>

# Human Adaptation, Plasticity and Learning for a New Sensory-Motor World in Virtual Reality

Michiteru Kitazaki

Department of Computer Science and Engineering,  
Toyohashi University of Technology, 1-1 Hibirigaoka, Tempakucho,  
Toyohashi, Aichi, 441-8580, Japan  
mich@cs.tut.ac.jp

**Abstract.** Human perception and action adaptively change depending on everyday experiences of or exposures to sensory information in changing environments. I aimed to know how our perception-action system adapts and changes in modified virtual-reality (VR) environments, and investigated visuo-motor adaptation of position constancy in a VR environment, visual and vestibular postural control after 7-day adaptation to modified sensory stimulation, and learning of event related cortical potential during motor imagery for application to a brain-machine interface. I found that human perception system, perception-action coordination system, and underlying neural system could change to adapt a new environment with considering quantitative sensory-motor relationship, reliability of information, and required learning with real-time feedback. These findings may contribute to develop an adaptive VR system in a future, which can change adaptively and cooperatively with human perceptual adaptation and neural plasticity.

**Keywords:** Adaptation, Plasticity, Position constancy, Galvanic vestibular stimulation, ERD/ERS.

## 1 Introduction

Environments are not static or constant. We are living in changing environments. Thus, our perception and action adaptively change depending on everyday experiences or exposures to sensory information in changing environments. Neural processing underlying basis of the perception and action seems also plastic and adaptive for new environments. My colleagues and I have investigated how our perception-action system adapts and changes in modified virtual-reality environments. In this paper, I describe three topics relating to adaptive change and plasticity of our perception and action, and neural learning of motor imagery.

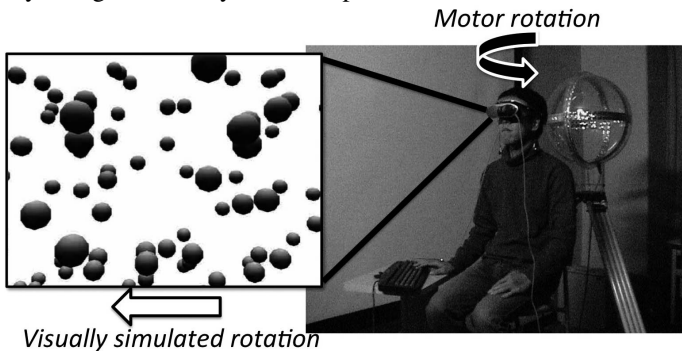
## 2 Adaptive Change of Position Constancy

Though our retinal image is always moving during head and body movements, we perceive a stable environment. This is called 'position constancy' or 'visual stability' in



perceptual psychology. Our brain has a compensation mechanism to stabilize the perceptual representation against motion of head and eyes [1]. It is well investigated whether the visual-motor system can be adaptively changed with an inter-sensory conflict situation. The most famous and traditional paradigm to investigate the adaptation of the visual-motor system is 'inverted vision' with a prism scope [2]. When one wears the prism scope, the perceptual world is inverted and he/she cannot help staggering around. After prolonged adaptation (1-4 weeks), the perceptual world gets back to proper orientation, and he/she becomes able to walk normally.

The experimental paradigm 'adaptation to a new visual-motor world' is a useful tool to investigate how the visual-motor system stabilizes our perceptual world [3]. We measured the position constancy during head turning in virtual reality environments to test its generality and selectivity [4-5]. Participants put on a head-mounted display (HMD) to observe a cloud of random spheres, which were stationary in the virtual environment. When participants turned their head rightward and leftward back and forth, the visual image presented on the HMD correspondingly moved by tracking head rotation with a Polhemus Fastrack sensor (Fig. 1). We varied the visual/motor gain for each trial. The visual/motor gain is 1.0 in the real world: when we rotate (yaw) our head rightward for 60 deg the retinal image moves leftward equivalent for 60 deg head yaw. In an experimental condition, we set the visual/motor gain 0.5, where the visual image on the HMD moved for 30 deg when participants moved their head for 60 deg. Participants observed the visual image during turning their head back and forth at the 0.5 visual/motor gain. Only after 2-min adaptation of the gain 0.5, the participants judged a smaller visual/motor gain than 1.0 as stable. This suggests that visual stability during active observation adaptively changes after only 2-min adaptation.

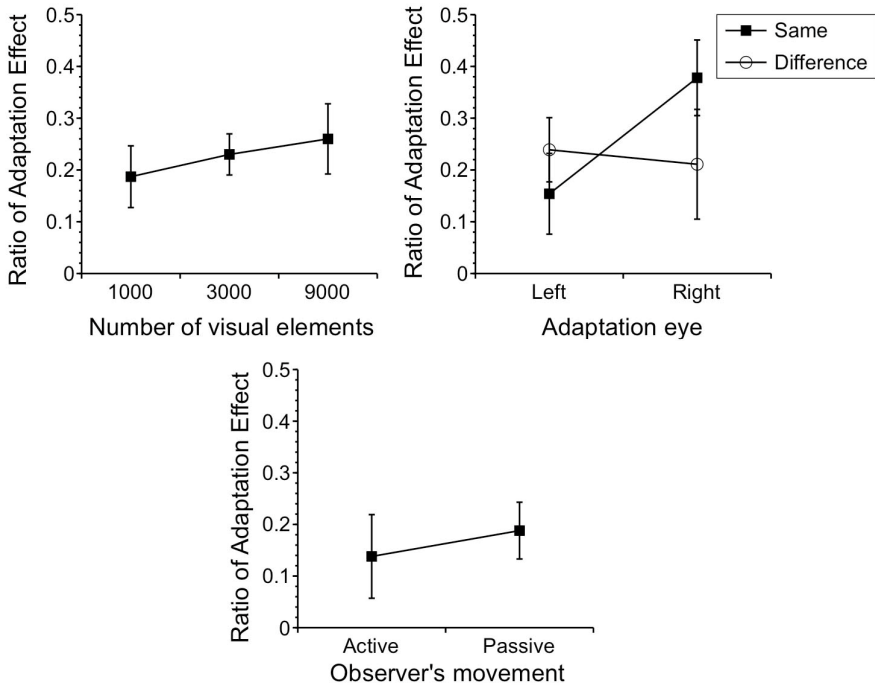


**Fig. 1.** Schematic of position-constancy experiment. Visual image changes depending on the head rotation. The visual/motor gain is ratio of visually simulated rotation by motor rotation.

This visual-motor adaptation occurred irrespective of amount of visual stimuli, active and passive head motions, and transferred between left and right eyes [4]. Thus, the visual-motor adaptation for visual stability quickly occurs and has high generality. However, the adaptation is partial, and only 13-37% of perfect adaptation (Fig. 2). It may due to our lifelong and robust adaptation to the real world of constant gain 1.0.

This visual-motor adaptation is good for surviving in changeable environments and using a virtual-reality system with limited spatio-temporal resolution. In subsequent

studies, we found that the visual-motor adaption for the position constancy was more effective if both the adaptation and test are performed on the left visual field than the other cases [5]. Thus, the left visual field appears to be more weighted to present visual information for effective adaptation than the other fields in a future adaptive virtual reality system.

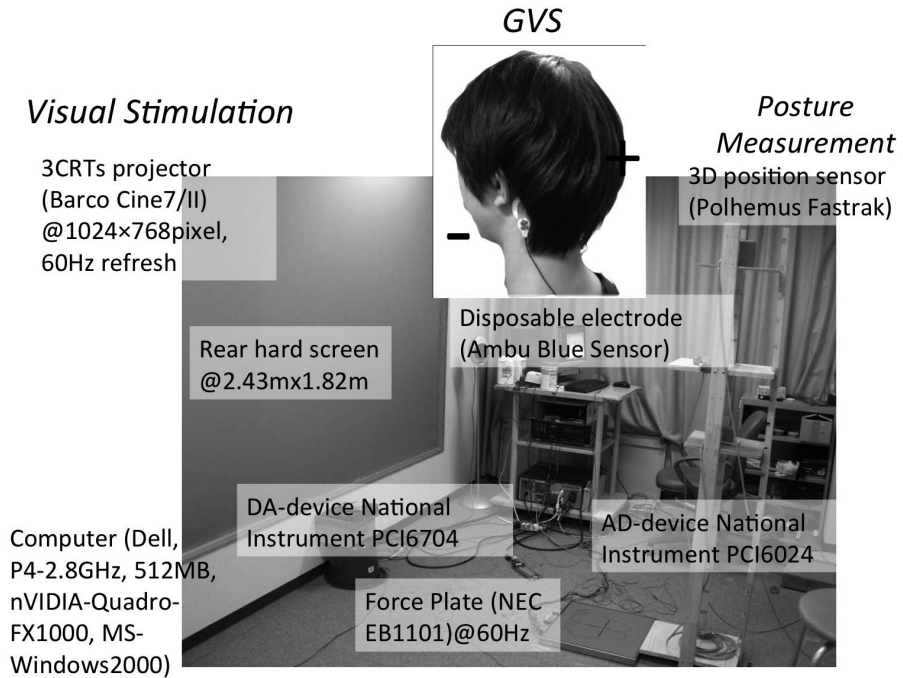


**Fig. 2.** Results of position-constancy experiments. Adaptation effects of visual-information richness (top left), effects of eyes (top right), and effects of active or passive movements (bottom) are shown. All conditions except for a condition of both adaptation and test with left-eye indicated significant adaptation effects ( $p < .05$ ). There were no significant main effects of the number of visual elements, adaptation eye, adaptation-test eye combination, or active-passive head motion ( $p > .05$ ).

### 3 Contributions of Vision and Vestibular Sense to Control of Posture

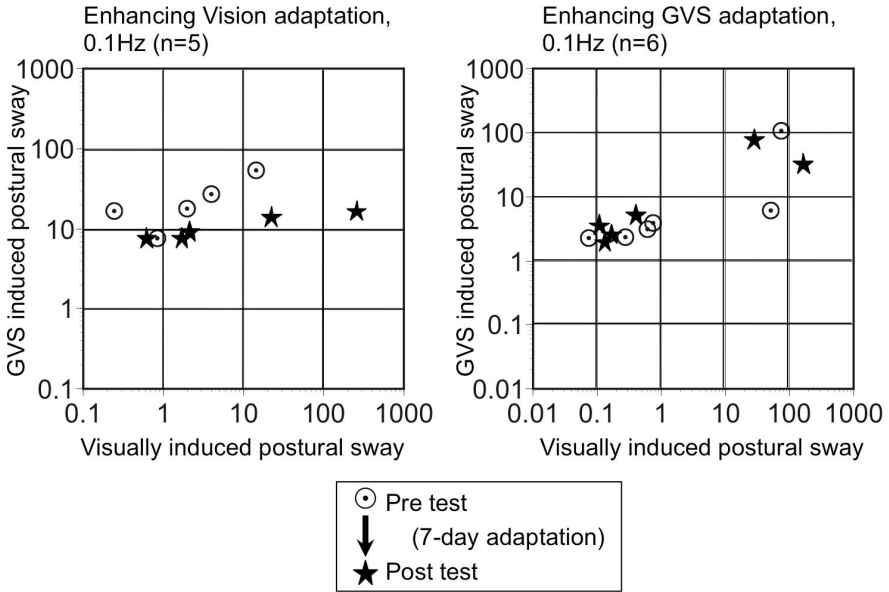
Human posture is controlled by multimodal process of visual, vestibular, and proprioceptive information. When a visual field contains a large visual motion, observers' body sway occurs at the identical frequency of the visual motion [6]. To investigate contribution of vestibular information to postural control, galvanic vestibular stimulation (GVS) is used. When a small current is applied to left and right mastoid processes, the observer inclines in the direction of anodal ear [7].

We measured postural sway induced by visual motion and galvanic vestibular stimulation (GVS) to investigate adaptive change of our visual and vestibular control of posture [8]. Participant's head position was monitored by a Polhemus sensor, and corresponding visual motion or GVS was continuously presented in real-time (Fig. 3).



**Fig. 3.** Apparatus of visual and vestibular postural sway experiments

Participants were divided into 4 groups: visual and GVS enhancing groups and visual and GVS inhibiting groups. For participants in visual or GVS enhancing groups, visual motion or GVS was presented to increase their voluntary sway (30 cm leftward and rightward at 0.2Hz). For participants in visual or GVS inhibiting groups, visual motion or GVS was presented to decrease their voluntary sway. After seven days adaptation (10 times of 1-min trial per day), participants in the visual and GVS enhancing groups showed more postural sway induced by visual motion and GVS, respectively (Fig. 4). However, we did not obtain equivalent results for the visual and GVS inhibiting groups. These results suggest that the long-term adaptation to enhancing action-yoked visual motion and GVS can modify weights on vision and vestibular senses to control posture.

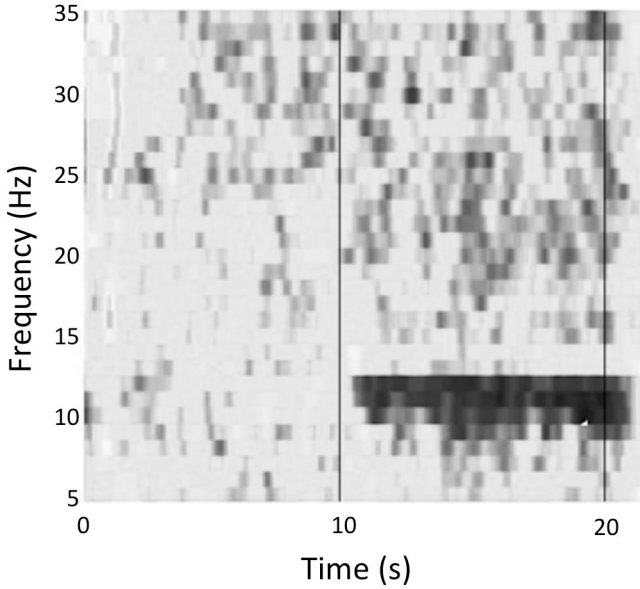


**Fig. 4.** Results of adaptation in visual and vestibular postural sway. Horizontal axis indicates sway power induced by visual motion, and vertical axis indicates sway power induced by GVS. Each mark indicates each subject's averaged result. Circles with a center dot are data of pre test, and stars are data of post test after 7-day adaptation.

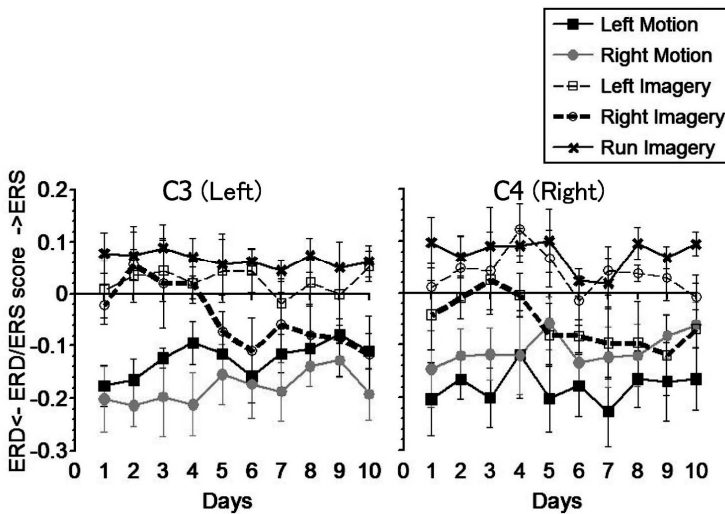
#### 4 Neural Learning of Motor Imagery

Finally, we measured brain activity during motor imagery. When human move their body such as fingers, event related de-synchronization (ERD) at approximately 8-13 Hz (mu-suppression) is observed at parietal lobe (motor cortex) by measuring electroencephalogram (EEG), and event related synchronization (ERS) is observed after ceasing movements. This ERD/ERS activity can be observed during motor imagery without actual movements (Fig. 5), and is used for brain-computer interfaces (BCI) [9-11]. Since the ERD/ERS occurs with silent reading [12], conversation without voice may be implemented to a BCI using ERD/ERS in future.

We made a real-time ERD/ERS feedback system for training and enhancing ERD/ERS induced by motor imagery [13]. When the averaged EEG power of C3 (center left channel) and C4 (center right channel) at 10-12 Hz was smaller than the power at the rest period before the feedback experiment (ERD), we presented a red bar whose length was increased upward corresponding to strength of ERD. When the EEG power was larger than the rest period (ERS), a blue bar was lengthened downward. The EEG power for ERD/ERS was calculated using 2s time-window data, and the visual feedback was updated at 10 Hz. Participants were asked to imagine clasp-ing their own left hand or right hand. They repeated 10 times of left-hand and right-hand motor imageries separately per a day with the visual feedback, and performed



**Fig. 5.** Example EEG spectrogram of ERD during motor imagery. Horizontal axis indicates time, and the subject was asked to image his hand movements (open and close his hand) during 10 - 20 s period. Vertical axis indicates frequency of EEG, and gray scale of data represents power calculated by short-time fft. Dark color indicates lower power and bright color indicates higher power than the rest period (0 - 10 s). ERD is found at 8-13 Hz during motor imagery period.



**Fig. 6.** Results of motor imagery learning. Horizontal axis indicates learned days. Vertical axis indicates ERD/ERS score. For both left and right hand motor imageries, the ERD at the contralateral channels (C3 for right hand, C4 for left hand) gradually increased as the learning progressed.

the training for nine days. We found contralateral increments of ERD during motor imagery as the training progressed (Fig. 6). These results suggest that human brain activity can be gradually changed or enhanced by the real-time visual feedback of brain activity contingent with motor imagery.

## 5 General Discussion

These three studies suggest that human perception system, perception-action coordination system, and underlying neural system could change for adapting to a new environment with considering quantitative sensory-motor relationship, reliability of information, and required learning with real-time feedback. A traditional aspect of virtual reality is that creating artificially most accurate sensory inputs is important to make virtual-reality systems. However, artificial engineering systems have limitation of spatio-temporal resolution and delays between inputs and outputs, thus it is difficult to perfectly mimic sensory inputs. A virtual-reality system can be designed adaptively by utilizing human perceptual and neural plasticity. The system may not require very accurate spatio-temporal resolution. It might be effectively implemented if both human perception-action system and virtual-reality system can change adaptively and cooperatively.

**Acknowledgments.** This research was partly supported by Grant-in-Aid for Scientific Research (B) #22300076 and Grant-in-Aid for Challenging Exploratory Research #23650060 from MEXT Japan.

## References

1. Wallach, H.: Perceiving a stable environment when one moves. *Annual Review of Psychology* 38, 1–27 (1987)
2. Stratton, G.M.: Some preliminary experiments on vision without inversion of the retinal image. *Psychological Review* 3, 611–617 (1896)
3. Wallach, H., Kravitz, J.H.: Rapid adaptation in the constancy of visual direction with active and passive rotation. *Psychonomic Science* 3, 165–166 (1965)
4. Kitazaki, M., Shimizu, A.: Visual-motor adaptation to stabilize perceptual world: Its generality and specificity. In: *Proceedings of 15th International Conference on Artificial Reality and Telexistence*, pp. 85–90 (2005)
5. Kitazaki, M.: Effects of retinal position on visuo-motor adaptation of visual stability in a virtual environment, i-Perception (in press)
6. Lestienne, F., Soechting, J., Berthoz, A.: Postural readjustments induced by linear motion of visual scenes. *Experimental Brain Research* 28, 363–384 (1977)
7. Day, B.L.: Galvanic vestibular stimulation: new uses for an old tool. *Journal of Physiology* 517, 631 (1999)
8. Kitazaki, M., Kimura, T.: Effects of long-term adaptation to sway-yoked visual motion and galvanic vestibular stimulation on visual and vestibular control of posture. *Presence: Teleoperators and Virtual Environments* 19(6), 544–556 (2010)
9. Pfurtscheller, G., Neuper, C.: Motor imagery activates primary sensorimotor area in man. *Neuroscience Letter* 239, 65–68 (1997)

10. McFarland, D.J., Miner, L.A., Vaughan, T.M., Wolpaw, J.R.: Mu and beta rhythm topographies during motor imagery and actual movement. *Brain Topography* 3, 177–186 (2000)
11. Birbaumer, N.: Brain-computer-interface research: Coming of age. *Clinical Neurophysiology* 117(3), 479–483 (2006)
12. Tamura, T., Gunji, A., Takeichi, H., Shigemasu, H., Inagaki, M., Kaga, M., Kitazaki, M.: Audio-vocal monitoring system revealed by mu-rhythm activity. *Frontiers in Psychology* 3, 225 (2012), doi:10.3389/fpsyg.2012.00225
13. Toyama, J., Ando, J., Kitazaki, M.: Event-related de-synchronization and synchronization (ERD/ERS) of EEG for controlling a brain-computer-interface driving simulator. In: *Proceedings of ACM Symposium on Virtual Reality Software and Technology*, vol. 16, pp. 239–240 (2009)

# An Asymmetric Bimanual Gestural Interface for Immersive Virtual Environments

Julien-Charles Lévesque<sup>1,2</sup>, Denis Laurendeau<sup>2</sup>, and Marielle Mokhtari<sup>1</sup>

<sup>1</sup> Defense Research and Development Canada Valcartier, Québec, Canada

<sup>2</sup> Computer Vision and Systems Laboratory, Université Laval, Québec, Canada

**Abstract.** In this paper, a 3D bimanual gestural interface using data gloves is presented. We build upon past contributions on gestural interfaces and bimanual interactions to create an efficient and intuitive gestural interface that can be used in a wide variety of immersive virtual environments. Based on real world bimanual interactions, the proposed interface uses the hands in an asymmetric style, with the left hand providing the mode of interaction and the right hand acting on a finer level of detail. To validate the efficiency of this interface design, a comparative study between the proposed two-handed interface and a one-handed variant was conducted on a group of right-handed users. The results of the experiment support the bimanual interface as more efficient than the unimanual one. It is expected that this interface and the conclusions drawn from the experiments will be useful as a guide for efficient design of future bimanual gestural interfaces.

## 1 Introduction

Gestural interfaces have attracted considerable attention in virtual reality (VR), since hands are used naturally by humans to interact with their environment. Bimanual interfaces, either gestural or not, have also received a significant amount of attention in recent years. It is now acknowledged that with carefully designed interactions, two hands can perform better than one on a given task [1,2,3].

This paper presents the design of a bimanual gestural interface for immersive virtual environments (IVEs). Gestures are chosen over physical props because the number of configurations and interactions possible with human hands are much greater. Many gestures convey meaning through culture and shared experience, and gestural interfaces can benefit from this [4]. Our goal is to exploit currently available technology for developing a two-handed *gestural* interface for IVEs. This paper provides the details as well as results of experiments that were conducted on the gestural interface, the preliminary version of which is described in [5].

The proposed interface was designed for IMAGE (shown in Figure 1), an application exploiting simulation and scientific visualization for improving user performance in understanding complex situations [6,7]. For the needs of this paper, IMAGE objects should be thought of as 3D objects which the user needs to manipulate and move around in the IVE.

In the proposed interface, hand gestures are captured by data gloves, rather than computer vision, for increased efficiency and reliability, but future work could study





**Fig. 1.** The proposed bimanual gestural interface in action

bare-hand gestural interfaces. This interface uses *static* gestures, i.e. static hand configurations or hand postures (e.g. clenched fist), not to be confused with *dynamic* gestures (e.g. waving hand).

To validate the efficiency of the presented interface, a comparative study was run on a group of users which were requested to perform a series of tasks using two variants of the interface; a one-handed variant, using only the dominant hand, and a two-handed variant. As a means of simplifying the task of designing and evaluating the interface, it is assumed that users are right-handed - similarly to the previous work surveyed in the next section. The time taken by users for performing the tasks in both variants was recorded. The results confirmed that the two-handed variant increased the performance and stability, thus supporting our bimanual interface design.

## 2 Previous Work

A major contribution in the domain of bimanual interactions is the work by Guiard [8]. By studying examples of real-world human bimanual interactions (e.g. writing on a sheet of paper), Guiard observed that work is split between the two hands in a structured way. He established three principles that describe asymmetric bimanual interactions and created a model that is known as the "kinematic chain model" for bimanual interactions.

The three principles established by Guiard are: 1) the right hand operates in a spatial frame of reference that is relative to the left hand, 2) the left and right hands operate on different spatial and temporal scales of motion and 3) the left hand usually precedes the right hand in action. Hinckley et al. have developed a bimanual interface for 3D visualization that uses Guiard's right-to-left reference principle with great efficiency [1]. Veit later validated these theories in a 3D environment [3].

Bimanual interfaces have also been shown on multiple occasions to be more efficient than their one-handed counterpart [9,3]. However, there are also cases where they end up decreasing performance due to poor interaction techniques and metaphors [10,11], so one must exercise caution in the design of two-handed interfaces and interaction metaphors.

On the topic of gestures, Nielsen et al. extensively covered the design and selection of gestures for an interface, providing ergonomical insight as to what type of gestures should be used in gestural interfaces [12]. In their bare-hands 3D user interface, Schlatmann et al. used the distance between hands for the selection and manipulation of objects [13]. Cabral et al. designed a vision-based gestural interface for 3D VR environments by segmenting a user's space into discrete zones (e.g., top-left, bottom-right, etc.) then mapped onto actions [14].







### 3 Gestural Interface Design

In this section and the following, we describe the proposed interface design, starting here with the manner in which hand gestures are used to control the environment. In this design, users perform static hand gestures to interact with the environment. It is based on Guiard's kinematic chain model, not in the way that objects are manipulated, but rather by the way interactions are initiated and completed by the hands. The following guidelines were used in the design of the interface:

- Guiard's principles for bimanual interactions are adopted.
- The right hand executes the interactions demanding the best precision.
- The number of gestures to be used is kept to a minimum, since there is a learning curve imposed on users for remembering the gesture set.
- Whenever possible, actions should be mapped to gestures that are semantically related. It is also important to give users proper instructions (cues as to why a given gesture was chosen) and training.
- The hand gestures used in the interface reuse some ergonomic guidelines found in [12], e.g. avoiding outer positions and relaxing muscles as much as possible.
- The beginning of interactions is associated with muscle tension and the end with the release of the tension (as proposed in [15]).

The left hand plays the role of a mode switcher and most of the direct manipulations are executed by the right hand. Hand gestures are executed by the left hand to select interaction modes, while the right hand performs the interaction itself, e.g. pointing, moving or scaling the desired object. The gesture set to be used in the environment is presented in Table 1.

**Table 1.** Hand gestures used in the environment

					
Clenched-fist	Thumb-up (or down)	Index pointing, thumb-up	Index pointing, thumb-down	Index and middle pointing, thumb-up	Pinch (Index, middle and thumb)

The fact that the left hand defines the mode of action means it will always act before the right hand, thus respecting Guiard's precedence principle. The left hand also acts on different spatial and temporal scales than the right hand (typically, once at the beginning and at the end of every interaction), following Guiard's second principle. The principle of right-to-left spatial reference should only be met if there were an interaction that is complex enough to justify its need, which is not the case yet for IMAGE. However, in a way, the right hand depends on the left hand to define its mode of operation so it could be claimed that the right-to-left reference principle is also met, although this is still open for discussion because it is not a spatial reference per se.

A support vector machine (SVM) is responsible for gesture recognition. SVMs provide a performance which is similar (both in execution time and accuracy [16]) to that of neural networks and do not require a network topology to be determined. The SVM can recognize a user's gestures from a pre-established set of gestures with which it was trained beforehand with good accuracy and precision for sets composed of around 10 gestures.

## 4 3D Interactions

The actions available to a user have been divided into three categories: 1) selection and designation of objects, 2) generic manipulations, which group all interactions for moving and positioning objects, and 3) system control, which represents all actions related to menus and to how the environment behaves. These interactions and their corresponding gestures are presented in Table 2.

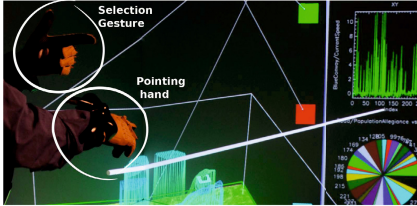
**Table 2.** Interactions and corresponding hand gestures. Refer to Table 1 for illustrations of the gestures

	Action	Left hand	Right hand
Selection	Designation and selection	Index pointing, thumb up	-
Generic manipulations	Move	Index pointing, thumb folded	-
	Rotate	Clenched fist	-
	Resize	Index and middle pointing, thumb up	-
System Control	Circular menus and numerical value modification	As needed, accept: Thumb up, cancel: Thumb down	Browsing: Pinch (thumb, index and middle)

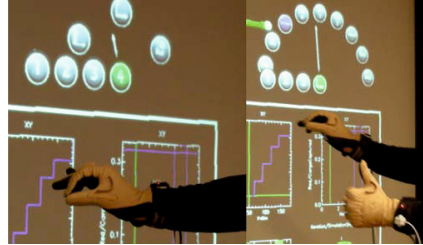
### 4.1 Designation and Selection

Designation is active - users need to maintain a gesture with their left hand for designation to be activated in the environment. As shown in Figure 2, designation is achieved by casting a ray with the user's right hand (the ray is displayed only once designation is activated), the first intersected object is then set as designated - the object is circumscribed

by a blue bounding box. Once designation ends (by switching to a different gesture), if an object was designated it then becomes selected. The use of an active designation should help the user focus on the data and information in front of him and not be distracted by the visual ray. As a visual feedback to the user, a selected object's bounding box is displayed in green. Once an object is selected, the user can then perform any of the other available actions on that specific object, with no need for reselection between interactions.



**Fig. 2.** Designation of an object with the left hand doing a gesture and the right hand pointing



**Fig. 3.** Circular menu used for system control. Left: user making his selection in the menu. Right: confirmation of a selection.

## 4.2 Generic Manipulations

Generic manipulations are relevant to positioning and orienting objects. As for designation and selection, these interactions are initiated by the left hand by performing the corresponding hand gesture, while the right hand controls the interaction:

- Moving an object: A selected object is moved using a metaphor that attaches it to a ray cast from the user's right hand. The distance from the hand to the object remains constant throughout the whole interaction and the object's orientation is adjusted so the object always faces the user.
- Rotating an object: For users to adjust all 3 degrees of freedom in rotation of a given object, an isomorphic rotation interaction was added which maps the user's right hand relative rotations (from the beginning of the interaction) on the selected object without scaling.
- Resizing an object: Objects can be resized using the right hand's distance to the object once the interaction is initiated. Bringing the hand closer to the body enlarges the object, while moving it towards the object makes it smaller.

## 4.3 System Control

System control is achieved through two different elements in the environment : control and circular menus (Figure 3).

Control menus are sets of icons representing actions - these icons can either be selected or dragged onto targets. A control menu is typically used for the creation of objects, while deletion of objects is achieved through a movement of the object above the head of a user, as if he wanted to throw it away behind him.

Circular menus are hierarchical menus that can be attached to objects for additional configuration capabilities. These menus expand once selected and the right hand's orientation can then be used to select a menu option. The user confirms or cancels a selection with his left hand by pointing his thumb up or down respectively. To limit fatigue, the thumb-down gesture must not point all the way down, only half-way (users are made aware of this). Symbolic input of numerical values is also provided with a behavior similar to that of circular menus, i.e., by rotating the right hand and accepting or cancelling with the left hand.

## 5 Experiments

In order to validate the presented design, the bimanual interface was compared to a unimanual variant similar in behaviour except that the single hand had to specify the mode of interaction and to control interactions at the same time (this corresponds to merging the last two columns of Table 2). For example, the one-handed selection of an object would imply making the 'index-pointing' gesture with the right hand, pointing the object with that same hand, and release the gesture while pointing the object. The experiments were conducted on a setup consisting of a wall screen with a stereoscopic projector and a pair of wired CyberGloves with Polhemus trackers for each hand.

The experiments consisted of a series of simple 3D manipulation tasks that reflect two purposes: first, they represent a normal use case of our application (strictly in terms of object selections and movements required), and second, they include enough repetition to assert whether or not a variant is better than the other. More precisely, the series of tasks consisted of 29 selections of objects (including one of a very small object), 10 movements, 7 drags (movement onto another object), 5 rotations, and one input of numerical value. One task was purposely harder than the others, calling for better precision - the selection of a very small object. The object was about 15 millimetres wide, and given the users distance to the screen (1 metre) the object was slightly less than one degree in width. An important point to reproduce these experiments is that the users must be compared with themselves, i.e. they must execute the experiments with both variants of the interface.

The users were given a training period before the actual experiment so they could become familiar with the various gestures and with the behaviour of the interface. Each user was timed while performing the task series twice, once using each variant of the interface. Since there is a learning curve for mastering the interface and the hand gestures, half of the users began with the unimanual interface to finish with the bimanual one, and the other half did the opposite. This crossover design limited the bias induced by the users' increasing expertise as they spent time using the interface.

Experiments lasted about one hour for each subject, with the calibration of CyberGloves (5 mins), a first training period (20 mins), followed by the evaluation of the first condition (10 mins), then followed by another shorter training period (10 mins) and the evaluation of the last condition (8 mins).

All participants were supervised during the experiments and notes were taken on any singularities that occurred during testing.

Two hypotheses were formulated for these experiments :

- **H1:** The bimanual variant should provide faster completion times than the unimanual one.
- **H2:** The bimanual variant should provide better precision of interaction and perform better than the one-handed variant for more difficult tasks.

This experiment aims to validate the presented bimanual gestural interface design as being efficient and reusable, while also providing additional information and insight for future work on the design of gestural interfaces. It should be noted that there are very few works comparing one-handed and two-handed *gestural* interfaces in the literature, so this is also a contribution of the paper.

Finally, all participants had good experience using computers, although not necessarily with 3D interactions or VR. A total of 14 participants completed the testing sessions, 7 starting with the unimanual variant and 7 with the other, providing a total of 28 datasets.

## 6 Results and Discussion

The collected data was split into four groups according to two factors; the first one being the starting variant, one-handed first or two-handed first and the second factor being the current number of hands used (either one or two).

The analysis starts by looking at the total experiment completion times for all users. Table 3 shows average completion times for every condition. The marginal means suggest that the number of hands had an effect on the performance of users while the starting variant did not.

The data also suggest that users improved progressively over time; i.e. that there was learning still taking place after training. A  $2 \times 2$  split-plot design was used to account for any differential carryover effect (it might be that users starting with the one-handed variant learned differently from users who started with the two-handed variant). The within-subjects factor is the number of hands and the between-subjects factor is the starting variant. This means that users are compared first with themselves inside their group (with the same starting variant), and then with the other group (with the other starting variant). To assert ANOVAs required assumption of variance homogeneity, we ran a Levenes test on the data, with a positive outcome ( $p = 0.95$ ).

**Table 3.** Experiment completion time mean values for each subgroup in minutes

	One-handed	Two-handed	Marginal mean
One-handed first	$10.76 \pm 1.11$	$6.58 \pm 1.29$	8.67
Two-handed first	$8.24 \pm 1.74$	$8.85 \pm 1.42$	8.54
Marginal mean	9.50	7.72	8.61

Table 4 contains the ANOVA table produced, where an  $\alpha$  level of 0.05 is used to assert statistical significance. Furthermore, a Student's *t*-test was performed on the first sub-group of each starting variant, ignoring the second set of measures per subject (i.e., only the data used to produce the first column of Table 3). This analysis is insensitive to any possible differential carryover effect. The null hypothesis is again rejected ( $t = 2.60$ ,  $df = 11.36$ ,  $p = 0.02$ ).

**Table 4.** ANOVA table for whole experiment completion times (split-plot design, with starting variant as the between factor and the current number of hands as the within factor)

Source	df	error	F	p	Significant?
Number of hands, NH (within-group)	1	152,654	17.291	0.0004	Yes
Starting variant, SV (between-groups)	1	137	0.016	0.9020	No
Interaction of NH $\times$ SV	1	65,826	7.456	0.0122	Yes

## 6.1 Sub-tasks Analysis

Two additional analyses were run on different sub-tasks to check if the outcome would be different than for the full experiment. Firstly, a single task (the dropping of one object onto another) was singled out from the experiment and analyzed. Secondly, a sub-group of tasks (four selections, movements and rotations) was taken out from the experiment, including task switching times, and was also analyzed.

For the single task, running the same ANOVA as before (mixed within-subjects and between-subjects) and looking at values returned for the number of hands suggested that there was again a difference, although it was not statistically significant with a  $p$ -value of 0.12.

The mixed within-subjects and between-subjects ANOVA ran on the group of tasks revealed that the two-handed variant became again significantly faster than the one-handed variant with a  $p$ -value of 0.040.

## 6.2 Precision Task

Not all participants were able to execute the hard task. It should be observed that 11 participants using the two-handed variant of the interface managed to select the small object, while only 3 succeeded with the one-handed interface. Although this does not reveal how more precise the bimanual interface is, it does validate H2 qualitatively.

## 6.3 Discussion

Looking back at Table 4, the data support the bimanual variant as significantly faster than one (H1 true). The interaction of the starting variant factor and number of hands factor proved to be statistically significant, responding to the hypothesis that there was

still learning taking place, although learning had a weaker impact than the number of hands. The starting variant was not found to provide a meaningful performance variance.

Both hypotheses were supported by the data, thus confirming our proposed gestural interface design as efficient. As an added benefit, the experiments also provide evidence that bimanual gestural interfaces operate significantly faster than their unimanual counterpart, and while this is not an entirely surprising result, it had never been tested before - at least to our knowledge. As a reminder, both interfaces were of equivalent strength, meaning that they had the same interactions available, used the same gestures and had the same expressive power.

While testing H1, it was also found that some tasks did not provide significantly different completion times for both variants of the interface while others did. The fact that the bimanual interface performed better on longer, more complex tasks, led us to formulate the hypothesis that the two hands in the proposed interface allow a user to think about his actions and chain interactions faster than the one-handed interface does. Such a hypothesis could be tested in future work by measuring switching times between interactions for one-handed and two-handed gestural interfaces.

The result from the selection of the small object support H2, hinting to the fact that the bimanual variant of the interface was more precise or allowed for greater stability. One potential reason for this is that the hand pointing and manipulating the objects does not have to switch gestures during the manipulation, allowing for increased steadiness. It was also stated informally by users during the experiments that the bimanual interface was less tiresome, possibly due to the fact that the workload was split between the hands.

## 7 Conclusion

In this paper, the design of a 3D gestural bimanual interface was presented. The goal was to create an efficient gestural interface to be used in IVEs by building upon past contributions on gestural interfaces and bimanual interactions. The interface uses the hands asymmetrically, assigning the mode switching to the left hand and the role of manipulation to the right hand.

A comparative study was run between the proposed interface and a one-handed variant on a group of right-handed users, who executed a series of tasks both unimanually and bimanually after a short training period. To accommodate for any bias or learning still taking place during testing, a crossover design was used with half of the users starting the experiment with the one-handed variant and the other half of the users starting with the two-handed variant. In the design of the experiments, the two hypotheses were tested on the IMAGE environment and were proven to be true.

The problem of bimanual gestural interactions was addressed on some specific points and there is room for further research, especially for collaborative work between the fields of VR, computer engineering and cognitive psychology. The conceived interface is an example of a state of the art 3D bimanual gestural interface design. It is hoped that it will help other researchers in the construction of future interfaces by providing a sound starting point as well as insight on bimanual gestural interfaces.



**Acknowledgements.** We gratefully acknowledge the Defence Research and Development Canada agency, MITACS-accelerate program and Thales for the funding and equipment which allowed us to conduct the research. We also wish to thank the numerous direct contributors to this project both from DRDC and the Computer Vision and Systems Laboratory (CVSL). Finally, we are grateful for the help provided by the students and research staff of the CVSL in the testing and evaluation experiments.

## References

1. Hinckley, K., Pausch, R., Proffitt, D., Kassell, N.: Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction* 5, 260–302 (1998)
2. Owen, R., Kurtenbach, G., Fitzmaurice, G., Baudel, T., Buxton, B.: When it gets more difficult, use both hands: exploring bimanual curve manipulation. In: *Graphics Interface 2005*, pp. 17–24 (2005)
3. Veit, M., Capobianco, A., Bechmann, D.: Consequence of two-handed manipulation on speed, precision and perception on spatial input task in 3D modelling applications. *Universal Comp. Science* 14, 3174–3187 (2008)
4. Mulder, A.G.: *Handgestures for hci*. NSERC Hand Centered Studies of Human Movement Project (1996)
5. Lévesque, J.C., Laurendeau, D., Mokhtari, M.: Bimanual gestural interface for virtual environments. In: *Proceedings of the IEEE VR Conference (VR 2011)*, pp. 223–224 (2011)
6. Lizotte, M., Bernier, F., Mokhtari, M., Boivin, E., DuCharme, M., Poussart, D.: IMAGE: Simulation for understanding complex situations and increasing future force agility. In: *Proceedings of the 26th Army Science Conference* (2008)
7. Mokhtari, M., Boivin, E., Laurendeau, D., Comtois, S., Ouellet, D., Levesque, J., Ouellet, E.: Complex situation understanding: An immersive concept development. In: *Proceedings of the 18th IEEE VR Conference*, pp. 229–230 (2011)
8. Guiard, Y.: Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Motor Behavior* 19, 486–517 (1987)
9. Ulinski, A., Wartell, Z., Goolkasian, P., Suma, E., Hodges, L.F.: Selection performance based on classes of bimanual actions. In: *3DUI 2009*, pp. 51–58 (2009)
10. Guimbretière, F., Martin, A., Winograd, T.: Benefits of merging command selection and direct manipulation. *ACM Transactions on Computer-Human Interaction* 12, 460–476 (2005)
11. Kabbash, P., Buxton, W., Sellen, A.: Two-handed input in a compound task. In: *SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, pp. 417–423 (1994)
12. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) *GW 2003*. LNCS (LNAI), vol. 2915, pp. 409–420. Springer, Heidelberg (2004)
13. Schlattmann, M., Klein, R.: Efficient bimanual symmetric 3D manipulation for markerless hand-tracking. In: *Proceedings of the Virtual Reality International Conference* (2009)
14. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the usability of gesture interfaces in virtual reality environments. In: *2005 Latin American Conference on HCI*, pp. 100–108 (2005)
15. Cerney, M., Vance, J.: *Gesture recognition in virtual environments: A review and framework for future development*. Iowa State University HCI Technical Report (2005)
16. Heumer, G., Amor, H.B., Weber, M., Jung, B.: Grasp recognition with uncalibrated data gloves - A comparison of classification methods. In: *2007 IEEE VR*, pp. 19–26 (2007)

# A New Approach for Indoor Navigation Using Semantic Webtechnologies and Augmented Reality

Tamás Matuszka, Gergő Gombos, and Attila Kiss

Department of Information Systems, Eötvös Loránd University, Budapest, Hungary  
{tomintt,ggombos,kiss}@inf.elte.hu

**Abstract.** Indoor navigation is an important research topic nowadays. The complexity of larger buildings, supermarkets, museums, etc. makes it necessary to use applications which can facilitate the orientation. While for outdoor navigation already exist tried and tested solutions, but few reliable ones are available for indoor navigation. In this paper we investigate the possible technologies for indoor navigation. Then, we present a general, cost effective system as a solution. This system uses the advantages of semantic web to store data and to compute the possible paths as well. Furthermore it uses Augmented Reality techniques and map view to provide interaction with the users. We made a prototype based on client-server architecture. The server runs in a cloud and provides the appropriate data to the client, which can be a smartphone or a tablet with Android operation system.

**Keywords:** Indoor Navigation, Augmented Reality, Semantic Web, Ontology, Mobile Application.

## 1 Introduction

Indoor navigation is one of the actively researched areas of nowadays. Good examples are the large buildings, complex supermarkets, warehouses, university campuses, museums, etc. where it takes a longer time to find a given destination. The importance of the research topic is illustrated by the increasing industrial interest. For example in the autumn of 2012 some large companies collaborated and set the aim to make a standard indoor navigation system.<sup>1</sup> For outdoor navigation there are tried and tested solutions but these methods cannot be applied for the indoor case. These systems are usually based on Global Positioning System (GPS) that requires permanent radio wave communication with satellites around the Earth. These radio wave signals cannot be provided within the buildings therefore this method is not working at indoor navigation. The following case illustrates well the complexity of the problem: the Ericsson had an indoor navigation research project but it was terminated in August 2012, thence the home page of the research is unavailable.<sup>2</sup>

---

<sup>1</sup>[http://www.computerworld.com/s/article/9230537/Nokia\\_Samsung\\_Sony\\_join\\_forces\\_to\\_improve\\_indoor\\_navigation](http://www.computerworld.com/s/article/9230537/Nokia_Samsung_Sony_join_forces_to_improve_indoor_navigation)

<sup>2</sup><https://labs.ericsson.com/apis/indoor-maps-and-positioning/>

Several attempts have been made to make accurate the indoor navigation. Existing methods use infrared signals [1], ultrasound [2], signal strength of various wireless connections such as GSM (Global System for Mobile Communications), Bluetooth, and Wi-Fi [3,4,5,6], inertial sensors to track user movements [7] as well as various digital image processing algorithms [8,9] to the positioning.

One of our objectives was to review the possibilities which are necessary for indoor navigation. The implemented system based on our research uses built-in sensors of mobile phone and Augmented Reality to provide the navigation. Both opportunities are based on the interaction of users. The system uses the advantages of semantic web to store the data and to compute the possible paths. The Semantic Web [10] aims at creating the “web of data”: a large distributed knowledge base, which contains the information of the World Wide Web in a format which is directly interpretable by computers. The goal of this web of linked data is to allow better, more sensible methods for information search, and knowledge inference. To achieve this, the Semantic Web provides a data model and its query language. The data model – called the Resource Description Framework (RDF) [11] – uses a simple conceptual description of the information: we represent our knowledge as statements in the form of subject-predicate-object (or entity-attribute-value). This way our data can be seen as a directed graph, where a statement is an edge labeled with the predicate, pointing from the subject’s node to the object’s node. The query language – called SPARQL [12] – formulates the queries as graph patterns, thus the query results can be calculated by matching the pattern against the data graph. The implemented prototype was tested in one of the campus of Eötvös Loránd University with the help of different types of users (students, teachers).

The structure of the paper is as follows. After the introductory Section 1 we present the related work in Section 2. Then, in Section 3, we give the details of the design of the implemented system. Then, we give two examples in Section 4 to demonstrate the usability of our application. Afterwards we describe our future plans in Section 5. Finally, we summarize our experiences in Section 6.

## 2 Related Work

There is a good summary of indoor navigation solutions in [13]. In this section we present some already existing indoor navigation methods. All methods have a common property that their main objective is similar to our goal: to develop an efficient indoor navigation system running on mobile phones (or on handheld PC).

Baus, Krüger and Wahlster present a hybrid navigation system in [14] that uses different technologies to determine the position of the user. They implemented a component to provide the indoor navigation. When the paper was written there were not available mobile phones with enough capacity, therefore authors used handheld PC-s.

According to Schmalstieg and Reitmayr [15] the data model has to be independent from any specific application and their implicit assumptions. The Semantic Web provides such a data model. In their paper they investigate how this model fits the requirements of Augmented Reality applications and how such a system can be

developed. The difference between our and their system is that they did not store maps and did not use map view for visualization.

Mulloni et al. [16] implemented a real time marker-recognizer tool on mobile phones. They compare solutions with the navigation method based on maps and with the navigation method based on GPS. Based on their tool they developed an application for indoor navigation. This was used in poster presentation of a conference. Similarly to our system they use the appropriate map and assign the markers to show the new location. The advantage of our system compared with Mulloni's system is that in our case the necessary information can be downloaded dynamically with a QR code and we do not need new installation while in their case it is required to download the new release of the software and to install it onto the mobile phone.

Mulloni, Seichter and Schmalstieg describe an Augmented Reality interface to support indoor navigation in [17]. They combine activity-based instructions with information points for the localization. We also used this technique. In addition the authors made comparative case studies to evaluate of their system.

### **3 Architecture of Our Prototype**

In this section we describe the main functions of the system. First, we review the investigated techniques, then present the architecture of our prototype and give the details of the functions.

#### **3.1 Investigated Techniques**

We studied several available navigation solutions. The first solution is a navigation based on WiFi signals that have been analyzed in a lot of papers [4,5,6]. In previous works the accuracy of positioning was increased by installing new devices (e.g. router). These devices are used as fingerprints. Our goal was to avoid the deployment of further WiFi devices, instead we wanted to use the already existing ones to guide the user to its destination.

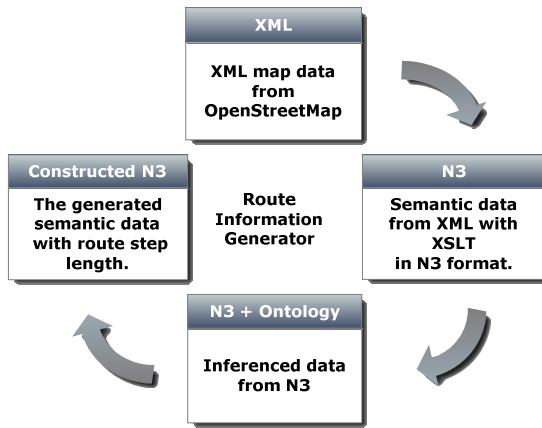
First we wanted to identify the exact position. The positioning was not correct, because of the measurement errors. Therefore we focused on defining large areas, so we split the map to cells. The signals were often unstable, therefore we often got wrong results. To solve this problem we created a solution to accept a cell only if it was adjacent to the previously correctly determined. Unfortunately, the tests showed that this still made a lot of mistakes. Later we added direction information to our measurement to correct the accuracy of the measurement [5], but it was not still enough to be able to determine the accurate position based on WiFi signals. Because of these unsuccessful attempts we discarded the WiFi navigation.

The next investigated solution was the pedometer that used the accelerometer and the compass of the mobile device [18]. We successfully apply this technique, the details are given in Section 3.3.

### 3.2 Server Components

The system has client-server architecture where the server is rarely used. The server is a cloud application using the CloudFoundry Micro system that provides a private cloud server for development. The application we made in CloudFoundry can be easily deployed to another cloud provider system.

The map necessary to the navigation was made by the OpenStreetMap map editor [19], which provides latitude and longitude coordinates to the corresponding points. The map stores POIs (Point of Interest), which are required for the final destination or the route calculation. This editor saves the data in XML format. We could transform this file automatically by means of the server to semantically interpreted content. We implemented the transformation with XSLT. The objects of the map are transformed to N3 format.



**Fig. 1.** The process of generating semantically annotated route data

The semantic layer contains well-annotated data which are extracted from the map and represent the environment. The correspondence between the map data and the semantic content was made with the INO ontology [20] that is extended by us. This extension contains the required concepts about navigation routes (e.g. POI, Passage, Corridor, Exit, etc.). In order to determine what points can be reached from a certain point, we needed the `hasAvailable` property that tells which points can be reached directly.

The map contained only those pieces of information that were necessary for drawing. We extracted the possible relation between two adjacent points from these pieces of information. Afterwards we ran an inference, which used the symmetric property to determine the possible steps. On the resulting routes a CONSTRUCT query provided by SPARQL was executed, which calculated the distance between two points based on the coordinates. The CONSTRUCT query makes a new RDF dataset from an existing RDF graph. This is the final dataset that is used by the client on the client side. This eliminates the need for storing all of the existing route information. Thanks to the `hasAvailable` property we can infer the possible routes from the points and their direct connections. This saves significant space for us.

### 3.3 Client Functionality

The client side is a device running Android operation system, which has built-in accelerometer, compass and camera. The first two are required to use the map function and the last is required to use navigation based on Augmented Reality. The system uses only those resources that are essential to the navigation, therefore it increases the energy efficiency.

First, we read the QR codes which store the current coordinates of the locations and the identifier of map located on highlighted places. The server sends all map data when starting the application so we do not need permanent internet connection to the navigation, therefore our solution is cost effective. After that the user chooses a destination and selects one from the possible visualization and starts the navigation.

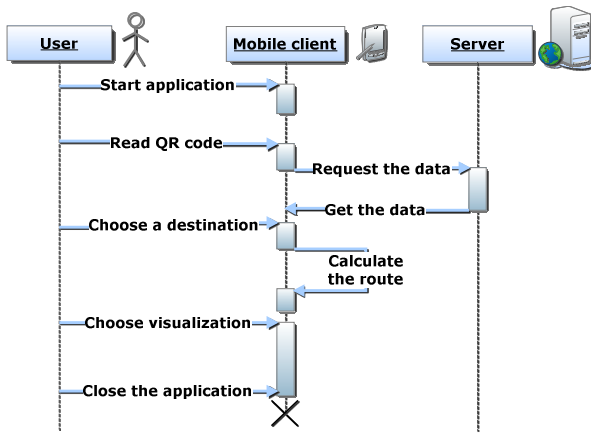
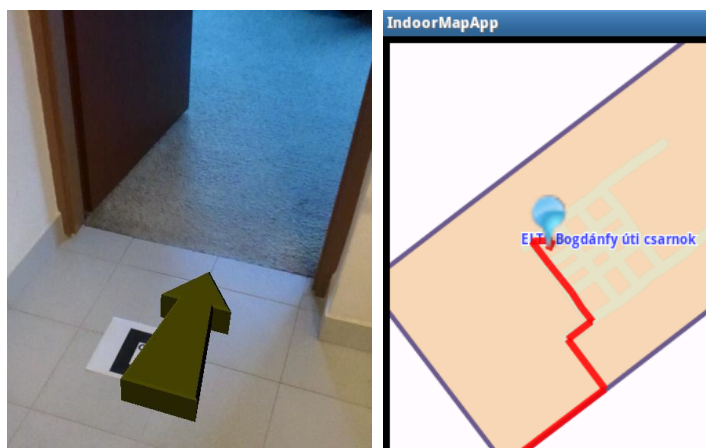


Fig. 2. The workflow of our system

One function of our system is the previously mentioned map mode with pedometer. It is based on the built-in accelerometer and compass. With the user interaction we can follow the motion of the phone. The phone displays the way that leads to the destination and the travelled distance on a map view. The pedometer has a calibration screen, where we can set our height. From the height the phone can calculate the length of the steps with mathematical methods. Since the step sizes may be different, it can cause small errors in positioning. A lot of small deviations lead to inaccuracies in the long term. In our case, the navigation is limited to a small area, so the error is less noticeable. On the basis of our experience, the pedometer method is able to determine our position with an accuracy of 50 cm. Our position can be updated with reading QR codes placed in the building.

Another feature of our navigation system is that we used Augmented Reality for the visualization [21]. Augmented Reality (AR) is a wide spread technology, by means of which the real physical environment can be extended by computer generated virtual elements. The system created this way is located between the real and the virtual world. We used the marker-based version of Augmented Reality, which uses

markers for registration in 3D. This technique uses the markers for reference points, where it displays the augmented content. We also investigated the natural feature tracking (NFT) option, but for two reasons we have decided not to use it. The first and most important point was that the application should also run on simple devices with small capacity. The cheaper types of Android phones do not have suitable hardware for NFT resource intensive operations. Another aspect was that the sign, which is used for navigation stands out from its environment. This aspect is better suited for traditional approach that uses markers.



**Fig. 3.** Visualization with AR and with pedometer on the map

Our idea was the following: we can use the crosses of map for navigation points and assign markers to them. A directed 3D arrow was defined to the marker. This arrow always points to the nearest point, which is in the route towards our final destination. Since the 3D models can be of large size, we used a model that produces the right direction with linear transformation. The detection of markers requires quadratic time. This depends on the number of the markers, because the system compares each marker to all others. In our case for the efficiency we can use only one marker for all. Because the markers are located also in POIs functioning as possible destinations thus the Augmented Reality based navigation provides accurate positioning.

## 4 Use Cases

The need is increasing for using indoor navigation in public places, office buildings, universities, hospitals, etc. In this section we present the functions of our application by using two scenarios namely an implemented and a fictional one. The first scenario is about the building of our research center and the second one is about a shopping center where the application can help the navigation.

Since the prototype is made for general purposes, of course, it is not limited to these two tasks, but with these two different use cases we want to demonstrate the flexibility of our application.

#### **4.1 Research Building**

In the first scenario we tried the system in our research building. We made a prototype application for this. The building has many rooms and a large lecture hall. Our aim was to help the navigation between important points (e.g. rooms, doors, tables) of the building. We mapped the building and each important point was saved to a map. On Fig. 3 one can see the map view and the shortest path from the current position to the exit on it. The application was tested on a device that has weak hardware (Samsung Galaxy Mini with 600 MHz processor) and running Android 2.2 operation system.

#### **4.2 Shopping Center**

The next example illustrates another possible case when our system could be used. Because the shopping centers usually have large area it is difficult and time-consuming task to find a given store inside them. The information boards are placed only in certain specific places, so the information is not available from anywhere inside the store. A lot of time can be saved by using a mobile device, which clearly shows the shortest path to our destination. For this purpose our application provides a cost effective solution that needs only cheap paper and plastic markers. These markers can be placed easily to the walls or the floor with stickers, so it is a really cheap way to help peoples to navigate. We can also use the map view based on the pedometer.

### **5 Future Work**

Several further developments are planned. One of the goals is to improve the pedometer method. The QR code can provide a solution for this problem. When the client recognizes the QR code then he gets the accurate position, so he can compare it with the position on the map view. This way it is possible to correct the position of pedometer.

A possible improvement of Augmented Reality view is to develop various navigation commands in addition to the arrows showing the right direction. With help of textual instructions we can learn how far away the destination is. As another opportunity the system does not only display an arrow but also shows the whole road section in front of the camera.

Currently to find the shortest path between two points a traditional graph search algorithm was used. We are planning to determine this path based on pure semantic web technology, namely we would obtain the shortest path with SPARQL queries. This method provides benefits at the personalization (when we would like to forbid the usage of certain path elements, e.g. staircase, elevator). Because of the inference ability of Semantic Web, the personalization can be done by using only one SPARQL query. For performance purposes the results of comparing of the two methods are also interesting for us.



## 6 Conclusion

Indoor navigation is one of the actively researched areas of nowadays. We achieved the following results. Firstly, we reviewed the state of the art of indoor navigation and investigated the technologies required to indoor navigation. A general, efficient system was designed based on the obtained results. Then we implemented a client running on Android operation system. With our system a user can navigate through an environment which has a map and contains arbitrary markers and QR codes. The application provides two different types of visualization for the navigation. Both are based on the interactions of users. The first is the pedometer method using the built in accelerometer and compass of the device. It also displays the way that leads to the destination and the travelled distance on a map view. The second visualization tool uses Augmented Reality, which extends the image of the camera with computer generated virtual objects. The system uses the advantages of Semantic Web to store the data and to compute the possible paths. Therefore our system combines two current technologies, Augmented Reality and Semantic Web to implement efficient and accurate indoor navigation. Our prototype was tested in one of the campuses of Eötvös Loránd University which was mapped and provided with markers.

## References

1. Want, R., Hopper, A., Falcão, V., Gibbons, J.: The active badge location system. *ACM Transactions on Information Systems* 10(1), 91–102 (1992)
2. Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., Hopper, A.: Implementing a sentient computing system. *Computer* 34(8), 50–56 (2001)
3. Otsason, V., Varshavsky, A., LaMarca, A., De Lara, E.: Accurate GSM indoor localization. In: Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H. (eds.) *UbiComp 2005*. LNCS, vol. 3660, pp. 141–158. Springer, Heidelberg (2005)
4. Zandbergen, P.A.: Comparison of WiFi positioning on two mobile devices. *Journal of Location Based Services* 6(1), 35–50 (2012)
5. Kessel, M., Werner, M.: SMARTPOS: Accurate and Precise Indoor Positioning on Mobile Phones. In: *Proceedings of the 1st International Conference on Mobile Services, Resources, and Users, Barcelona*, pp. 158–163 (2011)
6. Grossmann, U., Schauch, M., Hakobyan, S.: RSSI based WLAN indoor positioning with personal digital assistants. In: *Proceedings of the 4th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Dortmund*, pp. 653–656 (2007)
7. Merico, D., Bisiani, R.: Indoor navigation with minimal infrastructure. In: *4th Workshop on Positioning, Navigation and Communication*, pp. 141–144. IEEE Press, Milan (2007)
8. Hile, H., Borriello, G.: Information overlay for camera phones in indoor environments. In: Hightower, J., Schiele, B., Strang, T. (eds.) *LoCA 2007*. LNCS, vol. 4718, pp. 68–84. Springer, Heidelberg (2007)
9. Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., O. Gerl, Arnaudov, S., Lieberknecht, S.: An augmented reality museum guide. In: *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 103–106. IEEE Press, Cambridge (2008)

10. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* 284(5), 28–37 (2001)
11. Lassila, O., Swick, R.R.: Resource Description Framework (RDF) Schema Specification, <http://www.w3.org/TR/rdf-schema>
12. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>
13. Huang, H., Gartner, G.: A survey of mobile indoor navigation systems. In: Gartner, G., Ortogs, F. (eds.) *Cartography in Central and Eastern Europe*. LNG&C, pp. 305–319. Springer (2010)
14. Baus, J., Krüger, A., Wahlster, W.: A resource-adaptive mobile navigation system. In: *Proceedings of the 7th International Conference on Intelligent User Interfaces*, pp. 15–22. ACM, San Francisco (2002)
15. Schmalstieg, D., Reitmayr, G.: The world as a user interface: Augmented Reality for ubiquitous computing. In: Gartner, G., Cartwright, C., Peterson, M.P. (eds.) *Location Based Services and TeleCartography*. LNG&C, pp. 369–391. Springer (2007)
16. Mulloni, A., Wagner, D., Barakonyi, I., Schmalstieg, D.: Indoor Positioning and Navigation with Camera Phones. *IEEE Pervasive Computing* 8(2), 22–31 (2009)
17. Mulloni, A., Seichter, H., Schmalstieg, D.: Handheld Augmented Reality Indoor Navigation with Activity-based Instructions. In: *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 211–220. ACM, Stockholm (2011)
18. Link, J.Á.B., Smith, P., Viol, N., Wehrle, K.: FootPath: Accurate Map-based Indoor Navigation Using Smartphones. In: *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation*, Guimaraes, pp. 1–8 (2011)
19. Haklay, M., Weber, P.: Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7(4), 12–18 (2008)
20. Anagnostopoulos, C., Tsetsos, V., Kikiras, P.: OntoNav: A semantic indoor navigation system. In: *1st Workshop on Semantics in Mobile Environments*, Ayia Napa (2005)
21. Li, M., Mahnkopf, L., Kobbelt, L.: The Design of a Segway AR-Tactile Navigation System. In: Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., Krüger, A. (eds.) *Pervasive 2012*. LNCS, vol. 7319, pp. 161–178. Springer, Heidelberg (2012)

# Assessing Engagement in Simulation-Based Training Systems for Virtual Kinesic Cue Detection Training

Eric Ortiz<sup>1</sup>, Crystal Maraj<sup>1</sup>, Julie Salcedo<sup>1</sup>, Stephanie Lackey<sup>1</sup>, and Irwin Hudson<sup>2</sup>

<sup>1</sup> University of Central Florida, Institute for Simulation and Training, Orlando, FL  
{eortiz, cmaraj, jsalcedo, slackey}@ist.ucf.edu

<sup>2</sup> U.S. Army Research Laboratory – Human Research Engineering Directorate  
Simulation and Training Technology Center, Orlando, FL  
irwin.hudson@us.army.mil

**Abstract.** Combat Profiling techniques strengthen a Warfighter's ability to quickly react to situations within the operational environment based upon observable behavioral identifiers. One significant domain-specific skill researched is kinesics, or the study of body language. A Warfighter's ability to distinguish kinesic cues can greatly aid in the detection of possible threatening activities or individuals with harmful intent. This paper describes a research effort assessing the effectiveness of kinesic cue depiction within Simulation-Based Training (SBT) systems and the impact of engagement levels upon trainee performance. For this experiment, live training content served as the foundation for scenarios generated using Bohemia Interactive's Virtual Battlespace 2 (VBS2). Training content was presented on a standard desktop computer or within a physically immersive Virtual Environment (VE). Results suggest that the utilization of a highly immersive VE is not critical to achieve optimal performance during familiarization training of kinesic cue detection. While there was not a significant difference in engagement between conditions, the data showed evidence to suggest decreased levels of engagement by participants using the immersive VE. Further analysis revealed that temporal dissociation, which was significantly lower in the immersive VE condition, was a predictor of simulation engagement. In one respect, this indicates that standard desktop systems are suited for transitioning existing kinesic familiarization training content from the classroom to a personal computer. However, interpretation of the results requires operational context that suggests the capabilities of high-fidelity immersive VEs are not fully utilized by existing training methodologies. Thus, this research serves as an illustration of technology advancements compelling the SBT community to evolve training methods in order to fully benefit from emerging technologies.

**Keywords:** Kinesic cues, Engagement, Simulation-Based Training.

## 1 Introduction

Within the current tactical defense climate, Combat Profiling has put forth a critical Intelligence, Surveillance, and Reconnaissance skill set to assist the modern

Warfighter in threat detection [1]. Combat Profiling skills enhance a Warfighter's ability to maintain vigilance, situation awareness, and perceptual sensitivity of potentially threatening individuals within a combat environment. Combat Profiling training aids Warfighters in adopting a more proactive role akin to a hunter [2]. Rather than reactive post-incident tactics, Warfighters are trained to detect and assess pre-event indicators of potential threats by recognizing anomalies in the environmental and behavioral baselines, thereby, providing pre-incident, preventative, tactical planning.

Kinesics is the study of how nonverbal cues, body motion, and actions convey meaning [3]. In Combat Profiling, kinesics involves the ability to identify and analyze an individual's body language and affect [4]. Whether voluntary or involuntary, kinesic cues convey a great deal of information about an individual (i.e., attributes, motivation, attitude, and status) and the environment. These movements can indicate behavior that is atypical from the baseline and can allude to an individual's emotional state or pretense. Examples of kinesic cues include hand gestures, facial expressions, body language, and posturing.

Traditional Combat Profiling training methods utilize photographs and video footage for initial instruction in identifying behavioral cues of threats within the human terrain [1]. Live role players act out scenarios for experiential learning and profiling practice exercises. Although current methods are successful, limited availability of image and cinematic sources coupled with the high cost of hiring and training live role players restrict the cost effectiveness of widespread training [5]. Furthermore, Combat Profiling training is primarily conducted within the military, but principles of this training are applicable to other domains, such as homeland security and law enforcement.

Emerging research and development efforts have begun to investigate Virtual Environments (VEs) to enrich training of Combat Profiling skills. VEs offer a cost-effective and safe alternative to live environments [6]. An existing U.S. Navy research program is developing a comprehensive Combat Profiling training platform that transitions from computer-based training modules for declarative knowledge to an immersive team trainer for practical application of knowledge and skills [7]. Such efforts are making significant strides to improve the cost-effectiveness and deployability of virtual Combat Profiling training, but also prompt a need for experimentation to identify specific design requirements to improve the quality and effectiveness of virtual training systems. The VE literature indicates that effective virtual training may be affected by immersion and engagement [8]. To promote immersion and engagement, a goal of VEs is to "provide a compelling and effective medium for experiential, 'learn-by-doing'" opportunities [9]. Compelling VEs promote a "willing suspension of disbelief" that separates trainees from the real-world, enabling them to focus more intently on the training experience [10]. Although the body of research concerning simulation design factors that affect immersion and engagement continues to grow, there are still aspects that remain to be explored, refined, or translated to other training domains.

Visually representing Combat Profiling cues within Simulation-Based Training (SBT) systems requires investigation to support hardware and design requirements.

User immersion and engagement offer insight into developing threshold and objective requirements.

The experiment presented is one in a series investigating the role of immersive VEs and dynamic, high fidelity 3D virtual characters in deployable Combat Profiling training solutions. The use of kinesics or body language of virtual characters within a VE was empirically assessed to determine the effectiveness of virtual agent representations. The specific purpose of this research was to investigate the tradeoffs of training kinesic cues using a standard desktop or within a physically immersive VE system.

The following hypotheses were empirically assessed:

- $H_1$ =Participants will experience higher simulator sickness in the immersive VE condition.
- $H_2$ =Presence scores will be higher in the immersive VE condition.
- $H_3$ =Engagement scores will be higher in the immersive VE condition.
- $H_4$ =Technology acceptance subscale scores will be higher in the immersive VE condition.
- $H_5$ =Simulator sickness and technology acceptance subscale scores will be predictors of engagement.
- $H_6$ =Simulation engagement scores will be higher than pre-training engagement scores.

## **2 Method**

### **2.1 Participants**

Ninety students from the University of Central Florida's undergraduate population participated in this research experiment. Stipulations for participation included: U.S. Citizenship, age of at least 18 years old, and having normal or corrected to normal vision. Upon participation of the experiment, class credit was assigned accordingly.

### **2.2 Experimental Design**

This experiment investigated levels of immersion and engagement between two SBT configurations for training kinesic cue detection during Combat Profiling tasks. One configuration used a standard desktop system with a 22-inch display. The second configuration involved an immersive VE known as the Virtual Immersive Portable Environment (VIPE). The VIPE presents high-fidelity visuals on a 120-degree screen standing seven feet high within an enclosed space. Both configurations relied upon Virtual Battlespace 2 (VBS2), the U.S. Army's primary SBT platform. In order to maintain operational integrity, the experiment used VBS2 to supply virtual and constructive elements within the two hardware configurations studied. VBS2 provided tools to visually represent kinesic cues in an operationally relevant manner and the ability to develop customizable scenarios.

### 2.3 Kinesic Cue Detection

Kinesic cue detection training aims to enhance a trainee’s ability to identify kinesic cues such as hand and arm gestures, body language, and posture. Participants viewed pre-training content presented on PowerPoint slides. This included examples and descriptions of six kinesic cues—two cues per target affective state (Table 1).

**Table 1.** Kinesic cues displayed in experimental testbed

<b>Target Affective State</b>	Lying	Nervous	Aggressive
<b>Kinesic Cues</b>	Rubbing Neck Covering Mouth	Wringing Hands Check Six	Slapping Hands Clenched Fists

The mission environment (Figure 1) simulated a user walking on patrol with the task of identifying kinesic cues displayed and reporting each target’s affective state (i.e., nervous, lying, or aggressive). For the experimental conditions, three scenarios were developed to display kinesic cues including desert, suburban, and urban environments. All scenarios were created within VBS2 to emulate real world environments and included general features such as houses, buildings, foliage, people, animals, and vehicles. The desert environment included a non-geo-specific Middle Eastern scene with structures such as construction equipment, trucks, and trees. The suburban scenario consisted of parks, homes, and parked vehicles. The urban scenario reflects a non-geo-specific Middle Eastern setting including businesses, apartments, a playground, restaurants, produce stands, and an industrial area.



**Fig. 1.** Suburban mission environment displayed on both desktop and immersive VE condition

### 2.4 Measures

The following measures assessed participants’ feedback within the experiment. The Demographic Questionnaire gathers general biographical information from participants including age, gender, video-game experience, and computer competence. The Immersive Tendency Questionnaire is a measure used to determine

individual differences in the tendency to become deeply involved, or immersed, in activities [11]. The Simulator Sickness Questionnaire comprises of 16 symptoms designed to monitor participants' health status before and after exposure to a simulated environment [12]. The Presence Questionnaire comprises of 20 items that are related to the participant's perceived level of presence within each configuration [11]. The Engagement Measure is a subjective measure where participants rate their level of engagement [13]. This measure was administered once after the pre-training portion of the experiment (i.e., pre-training engagement) and once after exposure to the simulation environment (i.e., simulation engagement). The Technology Acceptance Measure is used to assess the participant's level of cognitive absorption, or engrossment, while using simulation technology [14]. Several subscales of the Technology Acceptance Measure address aspects of engagement including: temporal dissociation (i.e., unawareness of the passage of time), focused immersion (i.e., disregard for non-simulation distractions), heightened enjoyment, control, curiosity, perceived ease of use, and perceived usefulness.

## 2.5 Procedure

Upon arrival, the experimenters greeted the participants and each was randomly assigned to the desktop or immersive VE condition. Based on the condition, each participant was escorted by their experimenter to a designated lab area. At the location, the participant was asked to read the informed consent. Following this requirement, the participant was asked to complete the following questionnaires. These include: the Demographic Questionnaire, Immersive Tendency Questionnaire, and Simulator Sickness Questionnaire respectively. After completing the questionnaires, the participant was briefly instructed on how to complete the performance pre-test to follow. The performance pre-test required the participant to view a series of photographs demonstrating the kinesic cues addressed in this research area and attempt to identify the affective state of each cue. The participant then viewed the kinesic cue pre-training PowerPoint presentation. A Training Engagement measure followed the training slides. A five minute break was administered and upon conclusion the experimental condition began. Each participant completed a practice scenario for task familiarization followed by the experimental scenarios. The performance data was logged using an automated computer processing system.

There were three experimental scenarios that each lasted 15 minutes. Following each scenario, the participant completed the Simulator Sickness Questionnaire. After the final scenario, the participant completed the Presence Questionnaire, Technology Acceptance Measure, and the Simulation Engagement Measure. Final completion of the questionnaires was followed by a debriefing. The duration of the experiment was approximately two hours.

## 3 Results

An independent samples t-test was conducted to compare the immersive tendencies of participants randomly assigned to each condition. Results showed that there was no significant difference in the immersive tendency scores between groups suggesting

that the groups are representative of the same population. An additional independent samples t-test was conducted to compare the baseline simulator sickness of participants in each group revealing a significant difference between participants assigned to the desktop ( $M=5.56$ ,  $SD=8.98$ ) and immersive VE ( $M=10.45$ ,  $SD=15.49$ ) conditions;  $t(77)=-1.71$ ,  $p=0.027$ , 95% CI [-10.58, 0.80]. There was also a significant difference in the baseline simulator sickness subscale scores for disorientation and nausea, but not for oculomotor issues (Table 2).

**Table 2.** Results for baseline simulator sickness

Subscale	Desktop		Immersive VE		t(77)	p	95% Confidence Interval	
	M	SD	M	SD			Lower	Upper
Disorientation	3.23	9.88	8.75	17.56	-1.72	.005	-11.93	0.89
Nausea	3.31	5.22	7.00	12.65	-1.69	.001	-8.05	0.66
Oculomotor	7.05	11.07	11.08	14.63	-1.38	.258	-9.85	1.80

After exposure to the simulation environments, there was a significant difference in the disorientation subscale scores for the desert scenario in the desktop ( $M=7.51$ ,  $SD=17.41$ ) and immersive VE ( $M=19.55$ ,  $SD=32.12$ ) conditions;  $t(77)=-2.06$ ,  $p=0.042$ , 95% CI [-23.66, 0.96]. There was also a significant difference in the disorientation subscale scores for the urban scenario in the desktop ( $M=9.69$ ,  $SD=12.91$ ) and immersive VE ( $M=22.03$ ,  $SD=36.14$ ) conditions;  $t(77)=-2.01$ ,  $p=0.048$ , 95% CI [-24.55, -0.11]. There was no significant difference in the disorientation subscale scores between conditions for the suburban scenario. However, the descriptive statistics reveal a consistent trend with a lower mean disorientation subscale score for the desktop ( $M=10.77$ ,  $SD=17.73$ ) compared to the immersive VE ( $M=15.73$ ,  $SD=24.89$ ) condition. There was no significant difference in the nausea or oculomotor subscale scores between conditions. Likewise, the overall simulator sickness scores revealed no significant difference between conditions for all scenarios.

Separate independent samples t-tests were conducted to compare the perceived level of presence and the perceived level of engagement in the desktop and immersive VE simulation environments. There were no significant differences between conditions for the perceived level of presence or engagement. However, an independent samples t-test comparing the subscale scores of the Technology Acceptance Measure revealed there was a significant difference in the temporal dissociation subscale scores with higher scores in the desktop condition ( $M=10.72$ ,  $SD=3.87$ ) than in the immersive VE condition ( $M=8.45$ ,  $SD=3.62$ );  $t(77)=2.69$ ,  $p=0.009$ , 95% CI [0.59, 3.95]. A regression model was used to analyze average simulator sickness subscales (i.e., disorientation, nausea, and oculomotor) and temporal dissociation scores as possible predictors of engagement. The results showed that simulator sickness subscales were not a significant predictor of engagement. The temporal dissociation subscale significantly predicted engagement scores,  $\beta=0.41$ ,



$t(77)=3.97, p<.001$ . Temporal dissociation also explained a significant proportion of variance in engagement scores,  $R^2=0.17, F(1, 77)=15.72, p<.001$ .

Paired samples t-tests were conducted to compare the perceived level of engagement for the pre-training and the simulation for each condition. There was no significant difference between pre-training and simulation engagement scores in the desktop condition. Interestingly, there was a significant difference in engagement scores for the immersive VE condition with higher engagement scores in the pre-training ( $M=26.53, SD=4.75$ ) than in the simulation ( $M=25.25, SD=5.52$ );  $t(39)=2.94, p=0.006, 95\% CI [0.40, 2.15]$ . A regression model was used to analyze temporal dissociation scores as possible predictors of engagement in each condition. The temporal dissociation subscale scores significantly predicted engagement scores in the desktop condition,  $\beta=0.45, t(37)=3.06, p=0.004$ . Temporal dissociation also explained a significant proportion of variance in engagement scores in the desktop condition,  $R^2=0.20, F(1, 37)=9.38, p=.004$ . The temporal dissociation subscale scores significantly predicted engagement scores in the immersive VE condition,  $\beta=0.41, t(38)=2.79, p=.008$ . Temporal dissociation also explained a significant proportion of variance in simulation engagement scores in the immersive VE condition,  $R^2=0.17, F(1, 38)=7.79, p=.008$ . Spearman's rho correlations analyzed the correlation between simulation engagement scores and temporal dissociation overall and per condition. Across conditions, there was a positive, moderate correlation between simulation engagement and temporal dissociation,  $r_s(77)=0.091, p=0.002$ . Furthermore, there was a positive moderate correlation between simulation engagement and temporal dissociation in the desktop condition,  $r_s(38)=0.515, p=0.001$ . There was a weak positive correlation between simulation engagement and temporal dissociation in the immersive VE condition,  $r_s(39)=0.306, p=0.054$ . Overall, there were positive correlated relationships between simulation engagement and temporal dissociation within the desktop and immersive VE conditions.

## 4 Discussion

$H_1$  predicted that simulator sickness, presence, and engagement would be greater with the immersive VE than the desktop system. The immersive system was anticipated to cause more instances of simulator sickness because larger, immersive displays tend to cause episodes of disorientation, nausea, or oculomotor disruption [15]. Although the results for the disorientation subscale are consistent with expectations, the baseline difference between groups, with the immersive VE group's baseline significantly higher than the desktop group, may have skewed subsequent simulator sickness scores in the experimental scenarios. Contrary to the expectations of  $H_5$ , simulator sickness was not a predictor of engagement.

The results did not support the  $H_2, H_3,$  or  $H_4$  predictions that presence, engagement, and technology acceptance would be greater in the immersive VE condition. However,  $H_5$  was partially supported with the emergence of the temporal dissociation subscale on the Technology Acceptance Measure as a predictor of engagement. As suggested by results of the regression models, temporal dissociation, or the

unawareness of the passage of time, may indicate the level of engagement during a simulation experience. The correlation results suggest that as temporal dissociation increases, the level of simulation engagement also increases.

The results did not provide sufficient evidence for  $H_6$ , which predicted that simulation engagement would be greater than pre-training engagement. The level of engagement from pre-training to the simulation did not change in the desktop condition. However, engagement decreased significantly from pre-training to the simulation in the immersive VE condition. Perhaps, this decline in engagement was due to limitations of the experimental testbed design. In order to maintain consistency between conditions, only scenario events that appeared the same on both simulation displays were included. Pre-training content may have caused participants to anticipate a more compelling experience in the immersive simulation, but the scenario constraints for experimental consistency inhibited full utilization of the simulation environment capacity. Future experimentation may assess the effect of scenario variability on the level engagement.

Upon review of the results, it would appear that a desktop simulation system is more engaging than the immersive VE for kinesic cue detection training. However, it would be erroneous to accept such a conclusion without further consideration. There is a disparity in the desktop simulation's ability to simulate a peripheral view of the environment compared to the immersive VE. Perhaps, the forward focus of a flat panel display promotes greater engagement because all visual resources are allocated to the frontal view and not to the peripheral view. This is inconsistent with the operational environment where Warfighters' attention is divided among forward and peripheral lines of sight during patrol missions. Although a desktop simulator may inherently promote engagement, an immersive system, such as the VIPE, may provide more realistic opportunities for training Warfighters to practice observational and attentional strategies to overcome visual limits.

This experiment yields two design implications for SBT of kinesic cue detection with respect to increasing engagement. In his nine events of effective instruction, Gagné identified that instruction should begin with gaining attention and prompting learner expectancy [16-17]. Engagement may elicit attention and expectancy during exposure to new content. Therefore, the forward focused view of a desktop simulator may be appropriate for highly focused initial instruction and practice of observing and identifying kinesic cues in the environment. Once trainees master a basic understanding of the concept, a peripheral view provided by an immersive VE could offer a more realistic level of difficulty, challenging trainees to employ observational and attentional strategies for cue detection. Future experimentation may investigate how to leverage immersive VE system capabilities to train specific observational, attentional, and visual search skills and strategies.

Regardless of the simulation platform, a second implication is that engagement may be maintained by ensuring all phases of training (i.e., pre, during, and post) are designed to be equally compelling. In order to prevent a decline in engagement from one phase of training to the next, training expectations elicited in pre-training should be fulfilled through compelling practice scenarios in the during training phase. Although this experiment did not address post-training, the assumption of this

implication is that post-training activities should also include compelling elements or, perhaps, aspects that leverage the compelling features of the pre- and during training phases. This implication needs additional research to investigate strategies to maintain a consistent level of engagement throughout all phases of training.

## 5 Conclusion

This research paper compared engagement between SBT platforms for virtual kinesic cue detection training of Combat Profiling. Based upon the results, it is evident that software application is dependent upon the operational context and that the current training methods have not utilized such high-fidelity VEs for SBT. As such, research is needed to assess the capabilities of each platform and their ability to effectively train Warfighter's in detecting kinesic cues. Finally, developers of next-generation SBT systems need to consider how differing levels of engagement affect the Warfighter's ability to train effectively within a VE.

**Acknowledgement.** This research was sponsored by the U.S. Army Research Laboratory – Human Research Engineering Directorate Simulation and Training Center (ARL HRED STTC), in collaboration with the Institute for Simulation and Training at the University of Central Florida. This work is supported in part by ARL HRED STTC contract W91CRB08D0015. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL HRED STTC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

1. Gideons, C.D., Padilla, F.M., Lethin, C.: *Combat Hunter: The Training Continues*, pp. 79–84. *Marine Corps Gazette* (2008)
2. Hilburn, M.: *Combat Hunter: Experimental Marine Corps Project Aims to Turn the Hunted into the Hunter*. *Seapower*, pp. 60–62 (October 2007)
3. Birdwhistell, R.L.: *Introduction to Kinesics: An Annotated System for the Analysis of Body Motion and Gesture*. University of Louisville (1952)
4. Ross, W., Bencaz, N., Militello, L.: *Specification and Development of an Expert Model for “Combat Hunters.”* Technical report, Joint Research Laboratory Irregular Warfare Training USJFC (2010)
5. Frank, G.A., Helms, R., Voor, D.: *Determining the Right Mix of Live, Virtual, and Constructive Training*. In: *21st Interservice/Industry Training Systems and Education Conference*, pp. xx (2000)
6. Salas, E., Bowers, C.A., Rhodenizer, L.: *It is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training*. *Int. J. Aviation Psych.* 8(3), 197–208 (1998)

7. Schatz, S., Wray, R., Folsom-Kovarik, J.T., Nicholson, D.: Adaptive Perceptual Training in a Virtual Environment. In: Human Factors and Ergonomics Society Annual Meeting, pp. 2472–2476. Sage, San Diego (2012)
8. Cannon-Bowers, J., Bowers, C.: Synthetic Learning Environments: On Developing a Science of Simulation, Games, and Virtual Worlds for Training. In: Kozlowski, S.W.M., Salas, E. (eds.) *Learning, Training, and Development in Organizations*, pp. 229–262. Taylor & Francis, New York (2010)
9. Sims, E.M.: Reusable, Lifelike Virtual Humans for Mentoring and Role-playing. *Computers & Education* 49(1), 75–92 (2007)
10. Herrington, J., Oliver, R.: Patterns of Engagement in Authentic Online Learning Environments. *Australian J. Ed. Tech.* 19(1), 59–71 (2003)
11. Witmer, B., Singer, M.: Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7(3), 225–240 (1998)
12. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviation Psych.* 3(3), 203–220 (1993)
13. Charlton, J.P., Danforth, I.A.W.: Distinguishing Addiction and High Engagement in the Context of Online Game Playing. *Comp. Human Behavior* 23(3), 1531–1548 (2007)
14. Agarwal, R., Karahanna, E.: Time Flies When You’re Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly* 24(4), 665–694 (2012)
15. Lin, J., Duh, H., Parker, D., Abi-Rached, H., Furness, T.: Effects of View on Presence, Enjoyment, Memory, and Simulator Sickness in a Virtual Environment. In: *IEEE Virtual Reality* (2002)
16. Dick, W., Carey, L., Carey, J.O.: *The Systematic Design of Instruction*, 7th edn. Pearson, Columbus (2009)
17. Driscoll, M.P.: *Psychology of Learning for Instruction*, 2nd edn. Allyn & Bacon, Needham Heights (2000)

# Development of Knife-Shaped Interaction Device Providing Virtual Tactile Sensation

Azusa Toda<sup>1</sup>, Kazuki Tanaka<sup>2</sup>, Asako Kimura<sup>1</sup>,  
Fumihisa Shibata<sup>1</sup>, and Hideyuki Tamura<sup>1</sup>

<sup>1</sup> Graduate School of Information Science and Engineering, Ritsumeikan University

<sup>2</sup> Graduate School of Science and Engineering, Ritsumeikan University

1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577, Japan

toda@rm.is.ritsumei.ac.jp

**Abstract.** We have been developing “ToolDevice,” a set of devices to help novice users in performing various operations in a mixed reality (MR) space. ToolDevice imitates the familiar shapes, tactile sensation, and operational feedback sounds of hand tools that are used in everyday life. For example, we developed BrushDevice, KnifeDevice, TweezersDevice, and HammerDevice. Currently, KnifeDevice is insufficiency in force feedback. This paper proposes a tactile feedback model for cutting a virtual object utilizing two vibration motors and the principles of phantom sensation. We built a prototype to implement the proposed feedback model, and confirmed the usability of our model through an experiment. Finally, we redesigned KnifeDevice and implemented the tactile sensation on the basis of the results of the experiment.

**Keywords:** Mixed Reality, ToolDevice, phantom sensation, tactile sensation.

## 1 Introduction

We have been developing “ToolDevice” (Fig. 1), a set of devices to help novice users in performing various operations in a mixed reality (MR) space. ToolDevice imitates the familiar shapes, tactile sensations, and operational feedback sounds of hand tools that are used in everyday life.

In previous studies, we developed a handcrafting system [1][2] and a painting system [3][4] using ToolDevice. With TweezersDevice (Fig. 1c), used for picking up and moving virtual objects in the handcrafting system, we utilize a braking mechanism that uses a solenoid to provide force feedback when users pinch a virtual object. As for BrushDevice (Fig. 1a), used in the painting system, a reaction force mechanism helps a user to perceive tactile sensation when a user touch an object with the device [4]. These force feedback mechanisms are designed to be similar to that provided by the corresponding real world tools.

However, KnifeDevice (Fig. 1b), used for cutting virtual objects, had not exhibited such force feedback mechanism yet, because it is difficult to provide a reaction force of similar amplitude with that of a real knife operation by itself. Now, a user can cut virtual objects by placing them on a table and performing the cutting operation with

KnifeDevice. KnifeDevice already has a vibration motor to provide simple tactile feedback to show whether the device make contact with a virtual object. However, it is hard to feel and understand various virtual objects with different shape from their tactile feedbacks. An alternative tactile feedback whose mechanism can be built-in in a compact space is required.

This paper proposes a tactile feedback model for cutting a virtual object utilizing two vibration motors and the principles of phantom sensation. It provides similar sensation with that of real-world cutting by combining visual and tactile sensation. We also propose four methods to imitate real-world tactile sensations. We built a simple prototype implementing the proposed feedback model, conducted an experiment to compare the four proposed tactile sensation methods, and confirmed the usability of our model. Finally, we redesigned KnifeDevice and implemented the tactile sensation on the basis of the results of the experiment.



**Fig. 1.** (a) BrushDevice, (b) KnifeDevice, (c) TweezersDevice, (d) HammerDevice

## 2 Related Work

Tanaka *et al.* proposed a technique for representing forces acting on a virtual knife while it cuts a virtual object by using a haptic display called PHANToM to create tactile sensations [5]. PHANToM is a device that can provide haptic and tactile sensation in detail to the user by controlling the motion of the user's hand. However, PHANToM is required to be grounded, restricting the user's movements within the range of the mechanical linkages. Therefore, it is necessary to have a haptic display that provides much flexibility.

Kamuro *et al.* proposed Pen de Touch [6], which uses the non-binding force feedback mechanism. It provides force feedback for friction, reaction force, etc., thus enabling users to perceive the contact sensation and hardness of virtual objects. They developed a 3D modeling system [7] using Pen de Touch. Similarly, in this study, we also aim to develop a mechanism that provides non-binding force feedback.

Phantom sensation is known as a pseudo-tactile skin sensation perceived in an arbitrary place when two or more stimuli are provided simultaneously. The principles of phantom sensation were discovered by Von Békésy [8] and can be implemented using

vibration motors inside a device [9][10]. Using this method, a tactile feedback mechanism could be small and the user can move his/her hand freely. Using the principles of phantom sensation, we developed and implemented a tactile feedback model for KnifeDevice.

### 3 Tactile Sensation

#### 3.1 Analysis of Acting Forces While Cutting

In this study, we focus on a slicing action where a knife cuts an object not by pressing it but by moving through it. While cutting, forces are applied onto the knife from the user’s hand and the object being cut. The forces consist of horizontal friction force and vertical resistance force act on the knife. Vertical resistance force (Fig. 2) is the result of the object’s resistance and the vertical friction acting on the knife (lateral friction). Highly adhesive objects such as cheese and rice cakes have high lateral friction. On the other hand, non-adhesive objects such as wood and paper have low lateral friction. In this study, we assume that the objects being cut have low adhesivity; therefore, lateral friction could be negligible.

Fig. 3 shows the forces acting on the knife. When the knife is moved in the direction of movement,  $P$  is the force applied by the user, and  $F$  is the object’s resistance. Therefore, the following forces are being applied to the knife:  $F_n$  is the cutting force and  $F_f$  is the kinetic friction force.  $F_n$  and  $P_n$  act vertically on it. We define  $r_1$  as the distance between the fulcrum and the point of load and  $r_2$  as the distance between the fulcrum and the point of effort. As long as the knife does not rotate, the moment of forces is balanced, and the equilibrant is defined as follows:

$$r_1 F_n - r_2 P_n = 0 \tag{1}$$

To cut an object,  $P$  needs to be greater than  $F$ . As  $P$  increases, the movement becomes faster. However, this has no impact on the forces acting vertically. Therefore, we assume that  $P$  has no impact on the perceived feedback.

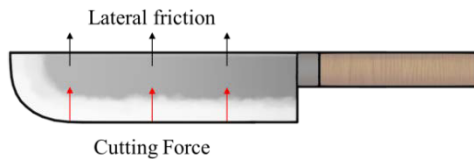


Fig. 2. Cutting force in vertical direction

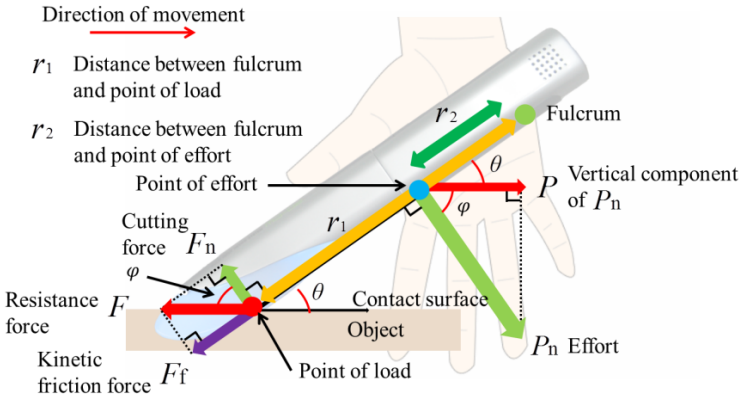


Fig. 3. Forces acting on KnifeDevice

### 3.2 Tactile Feedback Model

Because we use the same wooden material for all virtual objects in our handcrafting system, we can assume that  $F$  is always constant during the cutting operation. At this time,  $F_f$  does not change and  $F_n$  is defined by  $\theta$  (angle between KnifeDevice and surface of the virtual object) (Eq. 2) and  $\varphi$  (angle between the normal vector of KnifeDevice and surface of the virtual object) (Eq. 3).

$$F_n = F \cos \varphi \tag{2}$$

$$\varphi = \pi - \left( \theta + \frac{\pi}{2} \right) \quad \left( 0 \leq \theta \leq \frac{\pi}{2} \right) \tag{3}$$

During the cutting operation,  $P$  needs to be constantly greater than  $F$ ;  $P$  is calculated using Eqs. 1–3 and  $F$ . In this study, the minimum value of  $P$  (minimum force required to cut an object) is considered for tactile sensation. In other words, the maximum resistance force is used to represent tactile sensation.

When the contact point between the knife edge and object’s surface is fixed, the fulcrum point, point of effort, and point of load are static. At this time,  $F$ ,  $r_1$ , and  $r_2$  are constant, while  $P$  increases in proportion to  $\theta$ . On the other hand, when the contact point is changed and  $\theta$  is fixed,  $r_1$  changes depending on the point of load. At this time,  $F$ ,  $r_2$ , and  $\theta$  are constant, while  $P$  increases in proportion to  $r_1$ . On the basis of this observation, we consider  $\theta$  and  $r_1$  as the only factors that change  $P$ . In other words, we use these factors as parameters to control the vibration for presenting tactile sensation.

Vibration has several elements such as position, amplitude, and interval. However, with regard to interval, it is difficult to apply the duration of vibration or that of absence of vibration to our model. Therefore, we only change the amplitude and position of vibration while providing tactile sensation.



The position and amplitude in this context refer to the perceived position and amplitude of vibration when using the principles of phantom sensation, as described in chapter 2. To change the amplitude of the perceived vibration, the amplitudes of both the vibration motors are synchronized.

### 3.3 Prototype

We built a prototype (227 mm long) with two vibration motors (Linkman, 7AL09WA), one at each end of the device (Fig. 4). These vibration motors can be energized at 256 levels (from 0 to 255); the higher the voltage provided, the greater the amplitude of vibration. We conducted a preliminary study with three students in their twenties. The result showed that they did not perceive amplitudes less than level 30. They also did not accurately distinguish all amplitudes of vibration, so the levels were reduced to 16 (from 30 to 255 separated in steps of 15).

In Eq. 4,  $M_1$  and  $M_2$  are the amplitudes of the two vibration motors, and  $X$  is the position of the perceived vibration. When  $M_1$  and  $M_2$  are changed to 16 levels,  $X$  is also changed to 16 levels as follows:

$$M_1 = X \quad (0 \leq X \leq 15) \quad (4)$$

$$M_2 = 15 - X \quad (0 \leq X \leq 15) \quad (5)$$

### 3.4 Methods of Changing Pseudo-vibration

Pseudo-vibration is defined as the position of vibrations perceived on the basis of the principles of phantom sensation. We proposed the following four methods to change the position and amplitude of pseudo-vibration. These methods are combinations of angle and contact position used for changing the amplitude and perceived position of vibration.

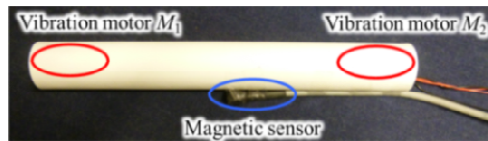


Fig. 4. Prototype

- (1) The amplitude of pseudo-vibration changes depending on the angle between KnifeDevice and the object's surface (angle  $\propto$  amplitude).
- (2) The perceived position of pseudo-vibration changes depending on the angle between KnifeDevice and the virtual object's surface (angle  $\propto$  perceived position).
- (3) The amplitude of pseudo-vibration changes depending on the position of contact between KnifeDevice and the virtual object's surface (contact position  $\propto$  amplitude).
- (4) The perceived position of pseudo-vibration changes depending on the position of contact between KnifeDevice and the virtual object's surface (contact position  $\propto$  perceived position).

In (1),  $P$  increases in proportion to  $\theta$ . Therefore, the amplitude is set to 0 when  $\theta$  is minimum, and the amplitude is set to 15 when  $\theta$  is maximum. In (2), the amplitude is set to 0 when the point of load is at the blade end. The amplitude is set to 15 when the point of load is at the front edge. In (3), as  $\theta$  increases, the force acting on the front edge increases. Therefore, pseudo-vibration is located at the front edge when  $\theta$  is maximum. The vibration is provided at the device end when  $\theta$  is minimum. In (4), pseudo-vibration is provided at the front edge when the point of load is at the front edge. The vibration is provided at the device end when the point of load is at the blade end.

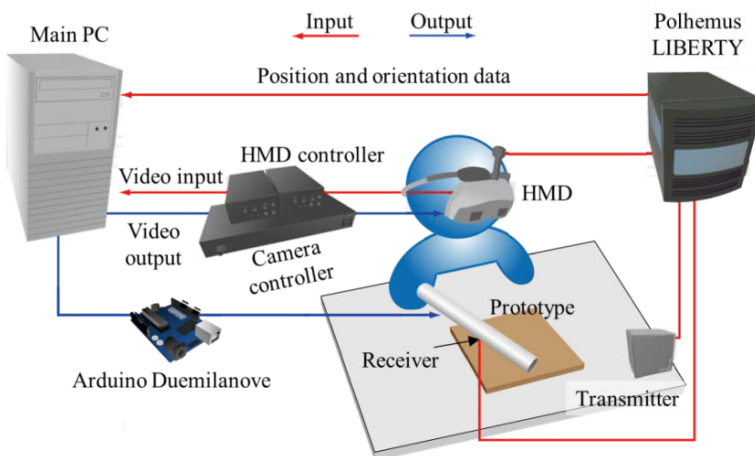
## 4 Experiment

### 4.1 Objective

We conducted an experiment to evaluate the usability of our proposed method. Specifically we analyzed the tactile sensation perceived when touching on a virtual object and slicing the virtual object with the knife. We also compared our proposed methods with the simple vibration method (the standard method) in which the amplitude of vibration is constant.

### 4.2 Environment

Fig. 5 shows the system configuration. We use a binocular see-through head mounted display (HMD; Canon VH-2002), which enables users to perceive depth. The HMD is connected to a video capture card (ViewCast Osprey-440) that captures input videos from the cameras built into the HMD. The position and orientation of the HMD and the device are tracked using Polhemus LIBERTY, a 6DOF tracking system equipped with magnetic sensors. A transmitter is also used as a reference point for the sensors.



**Fig. 5.** System architecture

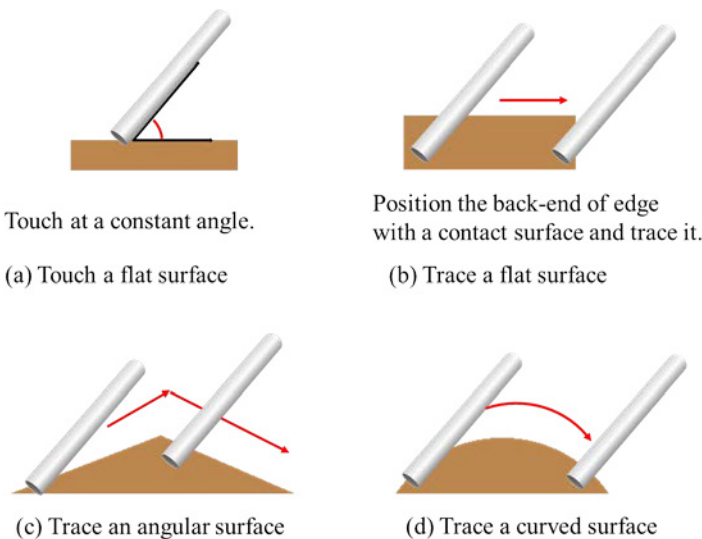
For creating an MR space, we first set the video captured by Osprey-440 as the background and then created a virtual viewing point in OpenGL by obtaining the position and orientation of the HMD from Polhemus LIBERTY. To control the vibration motors, we use the Arduino Duemilanove.

In the experiment, a red bar is rendered to indicate the front edge of the knife. When the knife comes into contact with the virtual object, a white sphere is rendered to indicate the contact point between the knife and the object.

### 4.3 Procedure

In this experiment, the subjects are required to perform the following four movements for the standard method and each of the aforementioned combinations (Fig. 6):

- (a) Touch a flat surface at 0, 45, and 90 deg
- (b) Trace a flat surface
- (c) Trace an angular surface
- (d) Trace a curved surface



**Fig. 6.** Movements used in the experiment

The order of the four methods used for each subject is randomized. In addition, between the trials for each method, we ask the subjects to try the standard method for comparing. After each trial, the subjects evaluated the four methods on a five-point scale (1: lowest, 5: highest) and compared them to the standard method that had been rated as 3. Five students in their twenties were the subjects.

## 4.4 Result and Discussion

The results are shown in Fig. 7. The bars indicate the average score for each method. From these results, we can conclude that almost all our proposed methods performed better than the standard method. Angle  $\propto$  amplitude (1) has the best score out of the four methods. Track an angular surface (c) has the best score for (a) to (d). The reason for this is that it was easy for the subjects to perceive an angular surface because the amplitude of the vibration increased when going through the angular part. On the other hand, the score of case (3)–(a) was lower than that of the standard method. One subject commented that he felt strange because the amplitude decreased as he pressed harder. As for the perceived position of the pseudo-vibration, two subjects did not perceive any change in position. They only perceived the change after they were told that the position could change. Thus, we conclude that the change in position might not be perceived without previous knowledge.

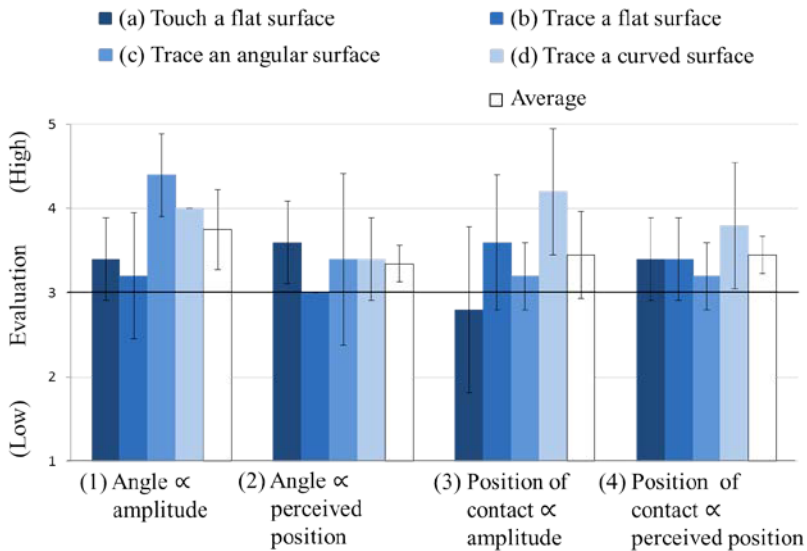


Fig. 7. Average score and standard variation

## 5 Implementation

### 5.1 Redesigned KnifeDevice

Fig. 8 shows the appearance and structure of redesigned KnifeDevice we developed. We confirmed the usability of our proposed method through our experiment and implemented it in KnifeDevice by mounting two large vibration motors at each end of the device. These motors have 256 levels of amplitude. In addition, to provide variety in the vibration in future, two smaller vibration motors were fitted into each end of the device.

Pressure-sensitive sensors are mounted with 256 levels (from 0 to 255) of sensitivity at the edge of the device and the gripper. By using these levels as input signals, users can turn on the tactile switch by applying a weak force on the table or cut objects in the air by gripping the device with a strong force.

## 5.2 User Study

As explained in chapter 4, angle  $\alpha$  amplitude (1) is the best method for providing pseudo-vibration, so we implemented it in redesigned KnifeDevice. Then, we conducted an experiment to confirm whether the tactile feedback is useful when it is implemented in KnifeDevice itself. The subjects, who were three students in their twenties, were required to cut various virtual objects, such as a cuboid, a hexagonal column, and a sphere, on a desk or in the air, and rate the usability of KnifeDevice. Our tactile feedback model was proven to be useful by this user study.

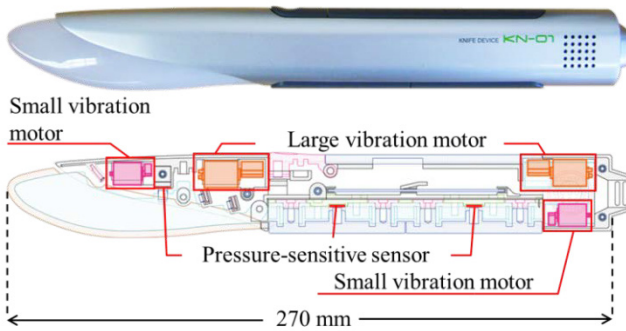


Fig. 8. Redesigned KnifeDevice

## 6 Conclusion and Future Work

In this paper, we proposed new methods to provide virtual tactile sensation while cutting a virtual object in the MR space. Our proposed methods utilize changes in the amplitude and position of vibrations in accordance with the angle between the device and virtual object's surface and the contact point between the device and object's surface. We implemented these methods in a simple prototype and conducted an experiment that compared the tactile sensations of the standard method in which the amplitude of vibration is constant against the four proposed methods. The results confirmed that almost of all our proposed methods provided better tactile sensations than those provided by the standard method. On the basis of the results of the experiment, we redesigned KnifeDevice. We conducted a user study and confirmed the usability of redesigned KnifeDevice. In future, we aim at improving tactile sensation by combining our proposed methods and using vibration motors of different intensities.

## References

1. Arisandi, R., Takami, Y., Otsuki, M., Kimura, A., Shibata, F., Tamura, H.: Enjoying virtual handcrafting with ToolDevice. In: Proc. UIST 2012, pp. 17–18 (2012)
2. Arisandi, R., Otsuki, M., Kimura, A., Shibata, F., Tamura, H.: Implementation of metal-working mode in mixed reality modeling system using ToolDevice. In: CD-ROM Proc. STSS 2012 (2012)
3. Otsuki, M., Sugihara, K., Kimura, A., Shibata, F., Tamura, H.: MAI Painting Brush: An interactive device that realizes the feeling of real painting. In: Proc. UIST 2010, pp. 97–100 (2010)
4. Sugihara, K., Otsuki, M., Kimura, A., Shibata, F., Tamura, H.: MAI Painting Brush++: Augmenting the feeling of painting with new visual and tactile feedback mechanisms. In: Adjunct Proc. UIST 2011, pp. 13–14 (2011)
5. Tanaka, A., Hirota, K., Kaneko, T.: Virtual cutting with force feedback. In: Proc. VRAIS 1998, pp. 71–75 (1998)
6. Kamuro, S., Minamizawa, K., Kawakami, N., Tachi, S.: Ungrounded kinesthetic pen for haptic interaction with virtual environments. In: Proc. IEEE Ro-Man 2009, pp. 436–441 (2009)
7. Kamuro, S., Minamizawa, K., Tachi, S.: 3D haptic modeling system using ungrounded pen-shaped kinesthetic display. In: Proc. IEEE VR 2011, pp. 217–218 (2011)
8. von Békésy, G.: Sensory Inhibition. Princeton University Press (1967)
9. Seo, J., Choi, S.: Initial study for creating linearly moving vibrotactile sensation on mobile device. Proc. IEEE 2010, 67–70 (2010)
10. Kim, Y., Lee, J., Kim, G.: Extending ‘Out of the Body’ saltation to 2D mobile tactile interaction. In: Proc. APCHI 2012, pp. 67–74 (2012)

# GUI Design Solution for a Monocular, See-through Head-Mounted Display Based on Users' Eye Movement Characteristics

Takahiro Uchiyama<sup>1,\*</sup>, Kazuhiro Tanuma<sup>1</sup>,  
Yusuke Fukuda<sup>2</sup>, and Miwa Nakanishi<sup>1</sup>

<sup>1</sup> Keio University, Yokohama, Japan

<sup>2</sup> Brother Industries, Ltd., Nagoya, Japan

takahiro.uchi@z8.keio.jp, kaz0414@gmail.com,

miwa\_nakanishi@ae.keio.ac.jp, yusuke.fukuda@brother.co.jp

**Abstract.** A monocular, see-through head-mounted display (HMD) enables users to view digital images superimposed on the real world. Because they are hands-free and see-through, HMDs are expected to be introduced in the industry as task support tools. In this study, we investigate how the characteristics of users' eye movements and work performance are affected by different brightness levels of images viewed with an HMD as the first step to establish a content design guideline for see-through HMDs. From the results, we propose specific cases based on the users' preferences for the brightness level of the image contents depending on the use of the HMD and the work environment. In one case, the users prefer low brightness levels, and in the other case, they prefer high brightness levels.

**Keywords:** Monocular, see-through head-mounted display, characteristics of users' eye movements, brightness of images.

## 1 Introduction

A monocular, see-through head-mounted display (HMD) enables users to view digital images superimposed on the real world. Because of their hands-free and see-through advantages, HMDs are expected to be introduced in the industry as task support tools. In fact, our previous research found that when workers performed wiring tasks by referring to a manual displayed by the HMD, human error decreased remarkably and task efficiency increased compared to using a paper manual [1]. While performance, size, weight, and resolution of the hardware have been improved, guidelines for the design of contents utilizing the see-through property have not been established. Therefore, in this study, we focus on the brightness of images, which is one of the most basic elements of the content and investigate the preferred content design with this type of HMD. In particular, we assume that wearable HMDs will be used

---

\* Corresponding author.

in a variety of applications and in various environmental conditions in future. Therefore, through experiments considering different types of usage and in different environments, we compared a situation in which we displayed high-brightness content all over with white background and black text, and another situation in which we displayed low-brightness content all over with black background and white text. We examined differences in users' visibility, fatigue, and eye movement characteristics.

In this study, we aim to reveal the preferences of users concerning the brightness of images displayed by HMDs depending on the application and environment.

## 2 Method

### 2.1 Experimental Tasks

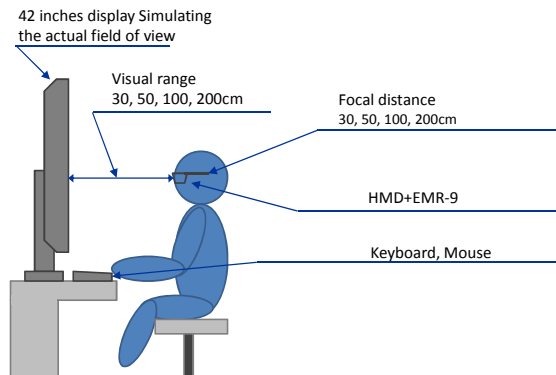
Because the see-through HMD is intended to be used in actual industry situations, we structured our experiments so that users referred to information on the HMD using two different patterns. In one case, users referred to information mainly by using the HMD (e.g., reading mail which is presented on the HMD). In the other case, users referred to both objects in the field of view and information on the HMD (e.g., comparing an operation order on the HMD with the actual work objects). Therefore, we provided the following two cases as the experimental tasks.

#### Task 1: No Interaction between HMD Information and the Real World

In this task, the subjects read only the text that was presented on the HMD. They wore an HMD (AiRScouter made by Brother; Figure 1) and sat in front of a large display (42 in, TH-42PX300, Panasonic), which simulated the actual field of view (Figure 2). The large display showed a full-color mosaic picture whose pattern changes every two seconds using 28 colors (Figure 3). The HMD displayed a variety of text words written continuously in katakana (Figure 4). The subjects were asked to find seven words of a specific type and indicate when they found them by pressing a key. The subjects repeated this task 10 times.



**Fig. 1.** Participant wearing an HMD



**Fig. 2.** Experimental Environment and Apparatus





i.e., the position in front of the half-mirror of the HMD that displays images. We measured the brightness of the large display that simulated the actual field of view. For each experimental environment, we measured the average value five times and averaged the results.

**Table 1.** Average brightness of each condition

	Visual range (cm)	The brightness of the black background (cd/m <sup>2</sup> )	The brightness of the white background (cd/m <sup>2</sup> )
Task 1 (Referring to HMD mainly)	30	24.30	64.30
	50	25.50	55.78
	100	25.60	53.73
	200	20.40	54.83
Task 2 (Referring to HMD and the actual world of view alternately)	30	18.63	50.20
	50	16.53	61.73
	100	17.57	46.23
	200	21.56	49.05

### 2.3 Measurements

We have summarized the steps followed for each task in Table 2.

**Table 2.** Measurements

	Task 1 (Referring to HMD mainly)	Task 2 (Referring to HMD and the actual world of view alternately)
Measurements	Work time: The time it took to read the content	Work time: The time taken for the comparison of one item
	Correct answer rate: Percentage of correctly detected the genre specified item	Correct answer rate: Percentage answered correctly match / mismatch of items
	Subjective assessment: <ul style="list-style-type: none"> <li>•0-100 point rating how comfortable doing tasks after each task</li> <li>•1-5 point rating eye fatigue subjects feel</li> </ul>	
	Eye movements: Measuring eye movements during the task with the eye-mark recorder (EMR-9 nac Image Technology) ※For convergence movement can be measured by measuring both eyes, we can obtain fixation point of the three-dimensional data Flicker value: Measured before and after each experiment with instruments flicker value (Type 2 501BTKK Takei Scientific Instruments Industry)	

### 2.4 Participants

The subjects were male and female adults from 18 to 24 years old. Twenty-four subjects performed the tasks using a white background, and 24 subjects used a black background.

### 2.5 Ethics

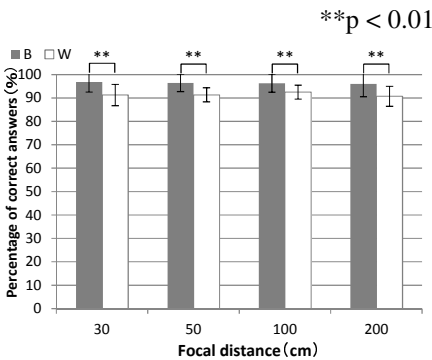
We obtained the informed consent of the participants.

## 3 Results

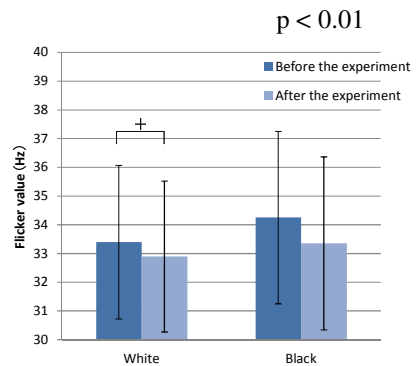
### 3.1 Task 1: No Interaction between HMD Information and the Real World

#### Task Performance

First, we compared the number of correct answers from subjects that performed tasks using a black background with the number of correct answers from subjects using a white background. The combinations of focal distances and visual ranges showed a higher percentage of correct answers from subjects using a black background than from those using a white background. Figure 7 shows the percentage of correct answers for each focal distance at 50 cm as the actual field of view. We focused on the differences in flicker values (which indicate psychological fatigue) before and after the experiment in order to explore the cause of these results. Figure 8 shows the differences between the flicker values before and after the experiment for tasks with white and black backgrounds. These results mean that mental fatigue increased when subjects used a white background. Therefore, we consider that in conditions with the white background (i.e., the content has high brightness), the concentration of the subjects fell because psychological fatigue of the subjects increased. As a result, the subjects' ability to accurately read the text declined.



**Fig. 7.** Comparison of correct answers using white and black backgrounds (50cm viewing distance)



**Fig. 8.** Comparison of flicker values before and after the experiment using white and black back-grounds

In addition, we compared the working times needed to perform tasks with a black background with the times needed to perform with a white background. As a result, if the viewing range was short, there were some cases for which the working times with

a black background were longer than those with a white background. From these results, we inferred that the see-through property was high with a black background, so mosaic images of the actual field of view obstructed the reading of the text on the HMD. However, there were no differences in 12 of the 16 combinations of viewing ranges and focal distances. Therefore, only when the viewing distance was short, reading was inhibited when using a black background; otherwise, we consider that there was no difference in the use of a black or white background.

### Eye Movements

Next, we analyzed the eye movements of the subjects. In the analysis, we designated left and right eye movements as the X axis, up and down movements as the Y axis, and movements toward and away from the face as the Z axis (depth measurement). In the analysis of the movement of the line of sight on the XY plane, we focused on gaze and saccade. We defined gaze on the basis of theory that eye velocity is 5deg/sec or less” [2, 3]. We established a standard based on viewing more than three consecutive frames at the same position, which also considered the frame rate of the analyzer (62.5 fps). In addition, we defined saccades as viewing at the same position less than one frame. On the other hand, in the analysis of the Z-axis direction, we used a diopter value (1/focal length) as the indicator. Figure 9 shows the time-series changes in the diopter value of a subject during a task.

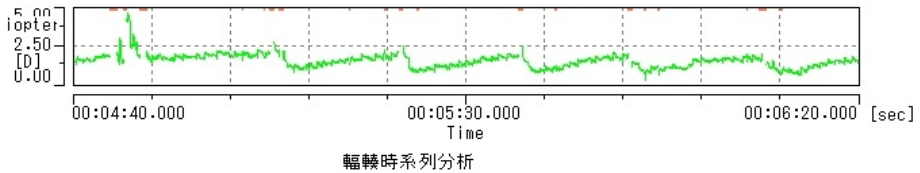


Fig. 9. Time-series change in diopter values

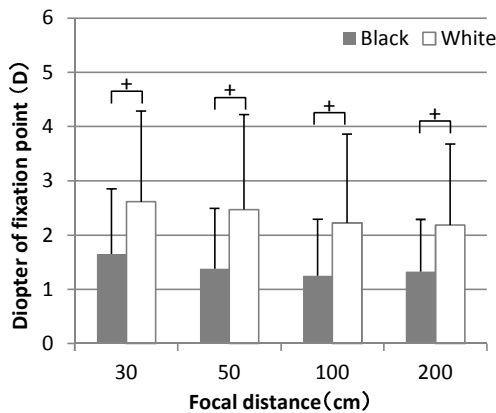


Fig. 10. Comparison of diopter values of the points of gaze with a white background and a black background (100 cm viewing distance)

We compared the average gaze time and average saccade distance during the tasks using a black background with those obtained using a white background, but there was no difference between them. On the other hand, we examined the diopter values when the subjects were gazing (diopter value of the point of gaze). The values with a white background were higher than those with a black background. Figure 10 shows the diopter values of the points of gaze when each focal distance to the actual field of view was 100 cm. This means that the subjects used a shorter focal length when viewing with a white background than when viewing with a black background. These results suggest that during the task of reading only the text on the HMD, the brightness of the image on the HMD affects the movements of the line of sight along the Z axis, although there is no effect on the movements of the line of sight on the XY plane. It is considered that this is related to an increase in the see-through property when the brightness of the images is low.

### **3.2 Task 2: Interaction between HMD and the Real World**

#### **Task Performance**

First, we found no differences when we compared the percentage of correct answers from subjects using a black background to those using a white background. When we compared the working times of subjects using a black background to those of subjects using a white background, we found that only when the viewing distance was 30 cm, the subjects using a white background usually had shorter working times than those using a black background. However, there were no differences in working times for other combinations of focal distances and viewing distances. These results indicate that when the subjects referred to both information on the HMD and objects in the actual field of view, the brightness of the HMD images did not significantly affect performance.

#### **Eye Movements**

Next, when we compared the gaze points on the XY plane of subjects using a black background to those of subjects using a white background, we observed two characteristics. One was a pattern of viewing by superimposing the images on the HMD on the actual field of view (Figure 11), and the other was a pattern of viewing in different positions without superimposition (Figure 12), when the subjects looked at the images on the HMD and the actual field of view alternately. Eight of the twenty-four subjects using a white background used a non-overlapping viewing pattern, while only two of the subjects using a black ground used this pattern. Therefore, it was revealed that there were more subjects using a superimposed pattern when viewing with a white background than those viewing with a black background taking advantage of the see-through quality.

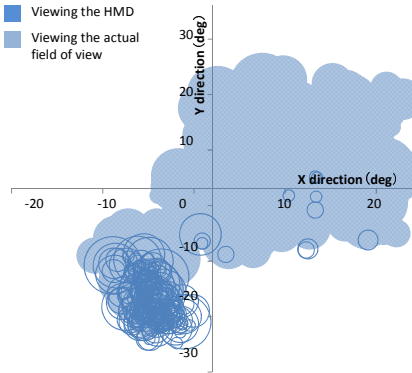


Fig. 11. Non-overlapping viewing

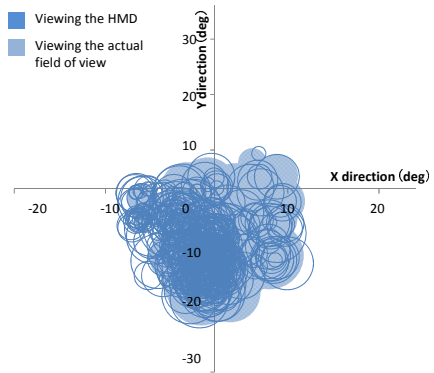


Fig. 12. Superimposed viewing

In addition, when we examined the movement of the line of sight along the Z axis, we found that there were roughly two features of the waveform of the time-series change in the diopter value. One was a sharp peak waveform (triangular wave in Figure 13), and the other was a flat peak waveform (rectangular wave in Figure 14). We found that there were many triangular waveforms when using either a white or black background with the same focal and viewing distances. However, when the viewing distance was longer than the focal distance of the HMD, there were many triangular waves when using a black background, but there were more rectangular waves than triangular waves when using a white background. When there were many triangular waves, the subjects focused on a close range for a short time for each instance. In contrast, when there were many rectangular waves, the subjects focused on a close range for a long time for each instance. The above experimental results show that the subjects took a shorter time to hold the line of sight when using a black background while referring to the text that was closer than the actual field of view on the HMD. It is considered that this is due to the fact that the see-through property is relatively low when using a white background, which requires the image brightness to be high. On the other hand, the see-through property is relatively high when using a black background, which allows the image brightness to be lower. Thus, in the former, there is a tendency for subjects to refer to information on the HMD and the actual field of view more distinctly, while in the latter, there is a tendency for subjects to refer to information on the HMD more superimposed on the actual field of view (Figure 15).

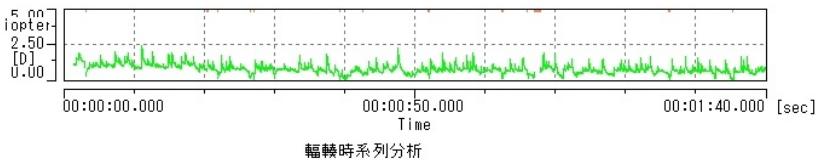


Fig. 13. Original waveform of diopter value (Example of the triangular wave)

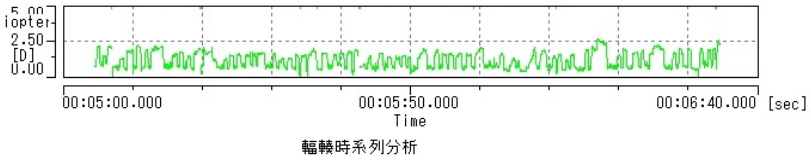


Fig. 14. Original waveform of diopter value (Example of the rectangular wave)

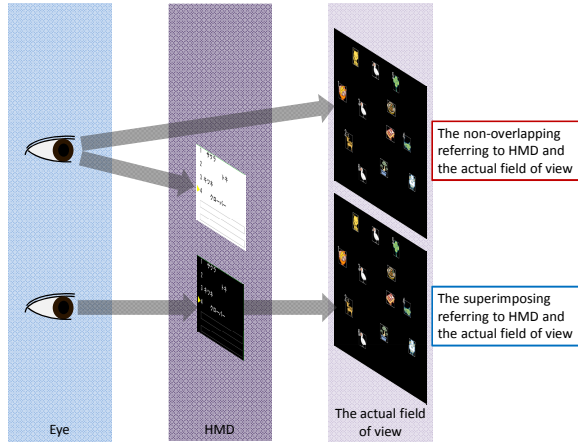


Fig. 15. Eye movements between the actual field of view and image on the HMD for black and white backgrounds

## 4 Discussion

First, for Task 1, with no interaction between the HMD and the real world (i.e., reading the information on the HMD at all times), if the user reads information on the HMD for a long time, mental fatigue is reduced when the image brightness on the HMD is low. However, if the brightness of the HMD images is low, the focus of the user moves away more easily because of the increase in the see-through characteristics. Therefore, the content might be difficult to read depending on the environment of the real field of view. Thus, when users view content with low brightness, they prefer white letters on a black background, for example, when using an HMD to read a newspaper or article in a taxi or train. However, when using the HMD outside in very bright or complex environments, increasing the brightness of images and reducing see-through properties would also be effective.

Second, for Task 2, with interaction between the HMD and the real world (i.e., watching both the information on the HMD and objects in the real field of view), if the brightness of the images on the HMD is high, the users watch the objects in the actual field of view and the information on the HMD separately in different positions. On the other hand, if the brightness of images on the HMD is low, the users have a strong tendency to watch the information on the HMD superimposed on the object in the actual field of view. Probably, this is also caused by the reduction in the

see-through property when the brightness of the HMD images is high. Thus, when viewing content with low brightness, users prefer white letters on a black background when it is important to get information from the HMD and the actual field of view at the same time, for example, when watching airplanes on the runway while viewing instruction in an HMD during the operation of an airplane. On the contrary, when viewing content with high brightness, users prefer black letters on a white background when it is important to get asynchronous information, for example, disrupting working and conveying such instructions while working.

## 5 Conclusion

In this study, we experimentally investigated the effect of differences in the brightness of HMD images on the characteristics of users' eye movements and work performance. From the results, we have proposed that when the brightness of the content is low, users prefer white text on a black background, and when the brightness of the content is high, they prefer black text on a white background depending on the use and environment. In future, on the basis of these suggestions, we will also consider more detailed design elements in the content and would like to connect our results to the establishment of a guideline for developing content design for see-through HMDs.

## References

1. Nakanishi, M., Okada, Y.: Development of an Instruction System with Augmented Reality Technology for Supporting Both Skilled and Unskilled Workers. *Human Factors in Japan* 11(1), 84–95 (2006)
2. Collewijn, H., Tamminga, E.P.: Human Smooth and Saccadic Eye Movements During Voluntary Pursuit of Differential Target Motions on Different Backgrounds. *J. Physiol.* 351, 217–250 (1984)
3. Yamada, M., Fukuda, T.: Definition of Gazing Point for Picture Analysis and Its Applications. *The Institute of Electronics, Information and Communication Engineers D J69-D(9)*, 1335–1342 (1986)



# Visual, Vibrotactile, and Force Feedback of Collisions in Virtual Environments: Effects on Performance, Mental Workload and Spatial Orientation

Bernhard Weber<sup>1</sup>, Mikel Sagardia<sup>1</sup>, Thomas Hulin<sup>1</sup>, and Carsten Preusche<sup>2</sup>

<sup>1</sup>German Aerospace Center, Institute of Robotics and Mechatronics, Wessling, Germany

<sup>2</sup>BMW Peugeot Citroën Electrification, Munich, Germany

{Bernhard.Weber, Mikel.Sagardia, Thomas.Hulin}@dlr.de,

Carsten.Preusche@bpc-electrification.com

**Abstract.** In a laboratory study with  $N = 42$  participants (thirty novices and twelve virtual reality (VR) specialists), we evaluated different variants of collision feedback in a virtual environment. Individuals had to perform several object manipulations (peg-in-hole, narrow passage) in a virtual assembly scenario with three different collision feedback modalities (visual vs. vibrotactile vs. force feedback) and two different task complexities (small vs. large peg or wide vs. narrow passage, respectively). The feedback modalities were evaluated in terms of assembly performance (completion time, movement precision) and subjective user ratings. Altogether, results indicate that high resolution force feedback provided by a robotic arm as input device is superior in terms of movement precision, mental workload, and spatial orientation compared to vibrotactile and visual feedback systems.

**Keywords:** Virtual environments, virtual prototyping, virtual assembly, haptic feedback, sensory substitution, usability, user study.

## 1 Introduction

While virtual reality (VR) technology is used in many fields of applications nowadays (like entertainment, edutainment, training and personal selection), our study focuses on virtual prototyping or assembly as it is one very promising approach to take advantage of the VR technology in the industrial domain. VR technology can be used to test mountability and usability in early design phases without physical prototypes [1], allowing significantly shorter product development cycles.

In the last years, virtual prototyping or assembly with so-called “digital mock-ups” (DMUs) is used routinely, for instance, in aviation or automotive industry. Virtual assembly has been defined as “the use of computer tools to make or ‘assist with’ assembly related engineering decisions through analysis, predictive models, visualization, and presentation of data without physical realization of the product or supporting processes” [2].

Thus, VR technology can be used to test and refine assemblability and evaluate changes of the assembly procedure [3]. Although approaches for automated assembly sequence planning (“computer aided assembly planning”, CAAP, e.g. [4]) exist, assemblers’ knowledge is still indispensable when trying to evaluate and optimize complex (dis-)assembly operations [5]. Moreover, VR technology also has the potential to serve as a training platform for future assembly workers.

The human-machine interface used to interact with the VR should allow for complex and natural manual interaction of virtual objects or tools. Furthermore, the VR system should provide sufficient sensory information to facilitate users’ spatial orientation and sense of immersion.

Immersive VR requires a high degree of visual realism, like high quality 3D visualization in real-time with unnoticeable delay. One major challenge, when simulating complex part interactions in VR settings is a realistic detection and display of collisions [6]. Collision information thus supports the human operator in understanding the spatial configuration, correcting position and orientation of the virtual object correspondingly and finding the target position [5]. In this work we mainly focus on three different modalities of displaying collisions in VR: visual, vibrotactile, and force feedback.

In the following section 2, we will provide a literature review on vibrotactile, visual, and force feedback of collisions in virtual environments. Next, methods (section 3) and results will be described (section 4) and discussed (section 5).

## 2 Collision Feedback in Virtual Environments

**Visual Feedback.** Visualization perhaps is the simplest and most frequently used form of collision feedback. Additional collision cues can be integrated easily, and no further output device is necessary. In prior studies, collisions have been visualized using arrows [7],[8], color changes [9],[10], or bar graphs indicating collision force and direction [11]. Nevertheless, one potential drawback of visualizing collisions might be that transferring visual as well as audio information into the force domain is cognitively demanding [11] and rather unintuitive. Moreover, adding dynamic visual aids to the VR visualization, which quickly change their shape, orientation, or color, potentially results in visual clutter and hence increased cognitive load. Compared to haptic feedback, however, visual feedback (e.g. symbolic arrows) has the potential to convey precise and unambiguous directional information how to solve an existing collision. In line with this notion, there is evidence that visual feedback is processed more rapidly and reaction times are shorter compared to haptic feedback [12].

**Vibrotactile Feedback.** One alternative to displaying collisions in VR is vibrotactile feedback. One general advantage of haptic feedback is that visual scenario information and haptic collision information complement each other as it is the case in real world scenarios [13]. In terms of information processing, using two different instead of one perception modality for conveying information should reduce the risk of overloading perceptual and cognitive resources. Compared to purely visual feedback, information can be presented independently from head or gaze direction [14].

Thus, users are able to plan object movements based on the visual scenario information and integrate collision information for trajectory corrections at the same time. Furthermore, in situations with obstructed view or occlusions, lacking visual information can partly be substituted by haptic information.

In contrast to force feedback systems, vibrotactile devices are less expensive, lighter, and provide larger workspaces. Besides, tactile feedback provides passive responses (i.e., no force is applied actively). Therefore, there is no conflict between feedback and the user's sense of position and less muscular fatigue [9]. Indeed, researchers could provide evidence that vibrotactile feedback can produce results similar to force feedback [14], [15] and even better results than visual feedback in teleoperation tasks [16].

**Force Feedback.** While vibrotactile feedback has the potential to improve the interaction in the VR, active force feedback systems significantly enrich the interaction in VR, resulting in a higher sensation of presence or immersion. The realistic and intuitive feedback of (collision) forces, significantly improves the user's performance when manipulating virtual objects (e.g. [8], [17], [18]). In his comprehensive work on force feedback in teleoperation, Massimino [14] gathered empirical evidence that force feedback is superior to auditory or vibrotactile feedback, when performing manipulation tasks or insertions with obstructed view. These performance benefits are mainly due to the fact that users are provided with realistic contact forces and are also forced into the correct orientation or position by force feedback. Thus, virtual objects can be guided more efficiently along kinesthetic constraints when there is not sufficient visual information (e.g. [8]).

### 3 Method

**Sample.** Thirty participants were recruited from the student and staff population of the German Aerospace Center. Moreover, twelve virtual assembly experts from automotive industry participated in the study, resulting in a sample of forty-two individuals ( $M_{AGE} = 30.3$  yrs.;  $Md_{AGE} = 27$ ). All participants read and signed a consent form.

#### Apparatus

*Visualization Hardware.* We used a 47" LCD monitor (200Hz) with 3D polarization display capability. Users sat approximately 1.5m away from the screen.

*Tracking System.* Users' hands were optically tracked using the Vicon Bonita system (240Hz) when testing visual and vibrotactile feedback conditions. Five infrared cameras pointing at the workspace were used, and the users had a tracked structure attached to their hands, consisting of four retro-reflective markers.

*The Vibrotactile Feedback Device (VibroTac).* VibroTac is a vibrotactile feedback device which was developed at the German Aerospace Center (see Fig. 1). It is used to apply vibrotactile stimuli to the human arm [19]. The device can be attached on a wide range of arm diameters while battery power and a wireless control interface contribute to unrestricted movement capability and user convenience.



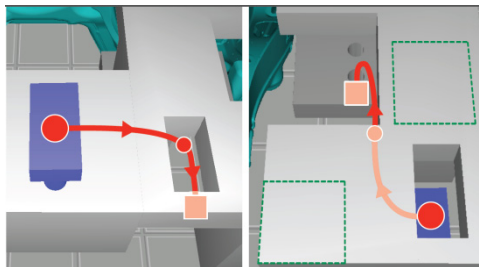
**Fig. 1.** The vibrotactile feedback device “VibroTac”

Six vibration segments are distributed around the human arm in equal distances. Several VibroTac devices can be worn for distributed feedback. The maximum data update rate is 1600 Hz.

*The Force Feedback System.* DLR’s haptic interface is composed of two light-weight robot arms which are attached horizontally at a column (see Fig. 3, right). The robot arms have a length of about one meter and the available workspace is similar to that of the human arm. The user’s hand is attached to system at a handle with Velcro fasteners. In order to minimize muscular fatigue, a feedforward algorithm is used to reduce the inertia of the robotic arms.

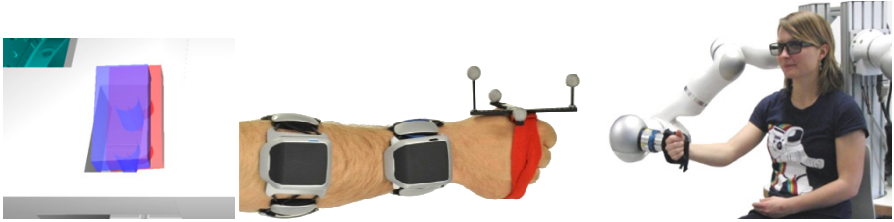
### 3.1 Experimental Task, Design and Procedure

**Experimental Task.** Participants started with a rectangular peg-in-hole task (see Fig. 2, left) with two different peg sizes (40 vs. 50 mm width, 100 mm length). Individuals had to move the peg from a pre-defined starting position (see Fig. 2, left, big red point) and pass the peg through a rectangular hole in the surface (52 mm width, 102 mm length).



**Fig. 2.** Peg-in-hole task (left) and narrow passage (right) with ideal movement paths (red); the surface area of the two columns are indicated with a green dashed line

Next, a more complex assembly task had to be performed with partially obstructed view (see Fig. 2, right). Subjects started below the rectangular passage and had to feed the rectangular peg (50 mm width) through two columns (80 vs. 60 mm distance). Finally, the two-pin object had to be assembled (see Fig. 2, right).



**Fig. 3.** Left: Visual collision feedback (red copy of controlled peg indicating collisions); Middle: VibroTacs and optical markers; Right: DLR's haptic interface

**Experimental Design.** A within-subjects design was utilized, i.e., each subject finished three experimental conditions (visual, vibrotactile, and force feedback). The order of the three conditions was randomized to control for potential time effects (like learning or fatigue).

**Procedure.** Participants were informed about the experimental task and procedure. A 3D model of the narrow passage structure was presented to the participants from various points of view and the path through the narrow passage was explained in detail. In the cases of visual and vibrotactile feedback, an elastic band with four optical markers for tracking the hand/ arm position was attached to subjects' dominant hand (see Fig. 3, middle).

*Feedback Conditions.* In the *visual feedback condition (V)*, collision between manipulated object and surface were indicated by displaying a red-coloured copy of the manipulated object which penetrated the virtual structures in case of collisions (see Fig. 3, left). Thus, the collision intensity was indicated by the distance between the manipulated object and its copy. To improve visibility even in case of occlusions, the controlled object and the copy were displayed transparently.

In the *vibrotactile feedback condition (VT)*, two VibroTacs were attached to the dominant forearm (see Fig. 3, middle).

The space between the devices symbolized the moved peg for the users, and the computed collision forces were accordingly mapped to this space. Participants were instructed that if there was a large-area lateral collision between manipulated object and surfaces, the corresponding tactors of both VibroTacs were activated. In case of a partial collision, for instance if there was a lateral offset between the manipulated object and a structure, the tactor(s) of the corresponding VibroTac would be activated. If there was a collision along the longitudinal axis (front or rear end) all tactors of the front or rear VibroTac were activated. The magnitude of the force was encoded in the amplitudes and signal frequency of the vibrating motors.

In the *force feedback condition (FF)*, participants manipulated the virtual objects by moving the haptic interface. In the present study, individuals' dominant hand was attached to a handhold with Velcro fasteners (i.e. only one robot arm was used). Contrary to the former conditions, user movements were recorded by the haptic interface.

In all conditions, individuals were told to avoid collisions when performing the task and to work as quickly as possible at the same time. In each experimental condition, individuals first completed the peg-in-hole task and the assembly path then.

In each task block, subjects completed a training trial first. Afterwards, subjects completed an experimental trial with the less difficult configuration first and then the more difficult configuration. Altogether, a number of 3 (1 training and 2 experimental trials) x 2 (task blocks) x 3 (feedback conditions) = 18 trials had to be completed.

After each feedback condition, subjects filled out the NASA-TLX questionnaire ([20]; German version), the System Usability Scale (SUS; [21]), and a questionnaire including items on spatial orientation, collision resolution and feedback clarity.

## 4 Results

Feedback modalities were evaluated using objective performance data and subjective user feedback in post-experimental questionnaires and interviews.

**Objective Data.** As objective performance indicators, we analyzed the time to complete the tasks (TTC) and the average collision forces during trials.

*Time-to-complete.* First, a repeated measures analysis of variance (ANOVA) with Feedback (V vs. VT vs. FF) and Difficulty (small vs. large peg) as repeated measures was performed on the TTC measure in the peg-in-hole task. While there was no significant main effect of Feedback ( $F(2;39) = 1.01$ ; *ns.*), a highly significant Difficulty main effect ( $F(2;40) = 27.82$ ;  $p < .001$ ) occurred. Furthermore, a significant two-way interaction between both factors ( $F(2;39) = 8.49$ ;  $p = .001$ ) was evident, with a significant Difficulty effect in the FF condition only. In this condition, completion times for the easy trials (small peg) were significantly lower ( $M = 7.71$ s, also see Tab. 1) than for the difficult trials with a large peg ( $M = 10.45$  s;  $t_{easy-diff.}(40) = 5.59$ ;  $p < .001$ ). Although no overall Feedback effect was evident, participants were fastest in the difficult trials when having VT collision feedback compared to the other feedback conditions ( $t_{VT-V}(40) = 1.89$ ;  $p = .07$ ;  $t_{VT-FF}(40) = 2.90$ ;  $p < .01$ ).

Similarly, analyzing TTCs in the narrow passage trials indicated no significant Feedback effect ( $F(2,39) = 0.57$ ; *ns.*), but a highly significant Difficulty effect ( $F(1,40) = 23.5$ ;  $p < .001$ ). In each Feedback condition the TTC was significantly lower in the easy trials compared to the difficult trials (all  $t_s(40) = 2.9$ ;  $p_s < .01$ ). No interaction effect was found ( $F(2,39) = .37$ ; *ns.*).

*Collision Force.* Analyzing the average collision forces in the peg-in-hole trials revealed significant Feedback ( $F(2, 39) = 23.6$ ;  $p < .001$ ) and Difficulty ( $F(1, 40) = 7.85$ ;  $p < 0.01$ ) main effects. A marginally significant interaction effect was found ( $F(2, 39) = 2.52$ ;  $p < .10$ ). Indeed, the Difficulty effect was largest and highly significant in the FF condition ( $t(40) = 3.93$ ;  $p < .001$ ), significant in the VT condition ( $t(40) = 2.67$ ;  $p < .05$ ) and non-significant in the V condition. Contrasting the Feedback conditions revealed highly significant differences between the FF condition ( $M_{easy} = 0.8$  N;  $M_{diff.} = 1.7$  N) and the V condition ( $M_{easy} = 8.1$  N;  $M_{diff.} = 11.3$  N; both  $t_s(40) > 4.8$  and  $p_s < .001$ ). Similarly, the average forces in the VT condition ( $M_{easy} = 10.4$  N;  $M_{diff.} = 13.8$  N) were significantly higher than in the FF condition (both  $t_s(40) > 3.2$  and  $p_s < .01$ ). No differences between VT and V were evident (both  $t_s < .88$ ).

Finally, we explored collision forces in the narrow passage trials. Again, highly significant main effects of Feedback ( $F(2;39) = 15.1$ ;  $p < .001$ ) and Difficulty

( $F(1;40) = 36.3; p < .001$ ) were evident. Moreover, the interaction effect of both factors was highly significant ( $F(2;39) = 14.9; p < .001$ ). Contrasting collision forces for the easy vs. difficult passage revealed a significant difficulty effect in the V ( $t(40) = 3.03; p < .05$ ) and a highly significant effect in the VT and FF condition (both  $t_s(40) > 4.3; p_s < .001$ ). Besides, forces in the FF condition were significantly lower for both the easy and the difficult passage (all  $t_s(40) > 3.87; p_s < .01$ ). The highest forces were measured in the VT condition in the difficult passage trials. In this case, forces were even significantly higher than in the V condition ( $t(40) > 2.3; p < .05$ ).

We did not find significant differences between the both subsamples (novices vs. VR experts) in the performance analyses reported above.

**Subjective Data. Workload.** A repeated measures ANOVA on the NASA-TLX overall score (scale ranging from 0-20) revealed a highly significant Feedback condition main effect ( $F(2, 40) = 8.6; p = .001$ ). We found significantly lower workload scores for the FF condition ( $M = 6.2; SD = 2.5$ ) compared to the V ( $M = 8.5; SD = 3.7; t(41) = 4.2; p < .001$ ) and the VT condition ( $M = 7.7; SD = 3.0; t(41) = 2.6; p = .01$ ). The average workload scores in the VT and V condition did not differ significantly ( $t(41) = 1.5, p = .13$ ). This result pattern was similar for the NASA-TLX items “Mental Demands”, “Performance”, “Effort”, and “Frustration”. No significant differences were found for the remaining items “Physical Demands” and “Temporal Demands”.

*Spatial Orientation.* [“I had a good overview of the spatial configuration, even in situations with restricted view or occlusions”, for all items scale ranged between 1-7; 1=“I fully disagree”; 7=“I fully agree”] A highly significant ANOVA main effect ( $F(2, 40) = 14.8; p < .001$ ) was found. Ratings in the FF condition ( $M = 5.2; SD = 1.2$ ) were significantly higher compared to the V ( $M = 3.3; SD = 1.6; t(42) = 6.0; p < .001$ ) and VT condition ( $M = 4.0; SD = 1.5; t(42) = 4.1; p < .001$ ). Moreover, ratings in the VT conditions were significantly higher than in the V condition ( $t(42) = 2.8; p < .01$ ).

*Collision Resolution.* [“There were repeated situations in which I did not know how to resolve a collision”]. A significant ANOVA main effect ( $F(2, 40) = 4.5; p < .05$ ) was also found. Individuals indicated that in FF conditions these situations occurred significantly least frequently in the FF condition ( $M = 2.8; SD = 1.6$ ) compared to the V ( $M = 3.6; SD = 1.6; t(42) = 2.3; p < .05$ ) and VT ( $M = 3.9; SD = 1.7; t(42) = 3.1; p < .01$ ) conditions. No such difference was evident comparing V and VT ( $t(42) = .90; ns$ ).

*Feedback clarity.* [“Collision feedback was unambiguous”] A highly significant ANOVA main effect ( $F(2, 39) = 33.9; p < .001$ ) was found. Ratings in the FF condition ( $M = 5.7; SD = 1.2$ ) were significantly higher compared to the V ( $M = 4.7; SD = 1.6; t(42) = 3.9; p < .001$ ) and VT condition ( $M = 3.4; SD = 1.4; t(41) = 8.0; p < .001$ ). Moreover, the difference between VT and V condition was highly significant ( $t(41) = 4.2; p < .001$ ).

*Usability.* ANOVA indicated a significant condition main effect ( $F(2, 41) = 12.9; p < .001$ ), with the highest usability rating in the FF condition ( $M = 82.2; SD = 12$ ; with a scale range from 0-100) and significantly lower ratings in the V ( $M = 72.6; SD = 17.1; t(42) = 3.7; p = .001$ ) and the VT condition ( $M = 68.6; SD = 15.2; t(42) = 4.9; p < .001$ ). The difference between the V and VT conditions did not reach the conventional level of significance ( $t(42) = 1.5; p = .15$ ).

**Table 1.** Performance and subjective measures: Means and standard deviations

<b>Objective Measures</b>	<b>Visual</b>	<b>Vibrotactile</b>	<b>Force Feedback</b>
<b>Time-to-Complete [s]</b>			
Peg-in-hole (Easy)	8.51 (4.91)	8.16 (5.21)	7.71 (3.51)
Peg-in-hole (Difficult)	9.79 (6.25)	8.40 (5.01)	10.45 (4.66)
Narrow Passage (Easy)	13.71 (7.09)	12.66 (8.06)	13.01 (6.1)
Narrow Passage (Difficult)	24.38 (28.2)	20.16 (20.8)	21.86 (15.29)
<b>Collision Force [N]</b>			
Peg-in-hole (Easy)	8.1 (9.70)	10.4 (19.1)	0.8 (0.9)
Peg-in-hole (Difficult)	11.3 (12.8)	13.8 (23.2)	1.7 (1.7)
Narrow Passage (Easy)	18.7 (29.4)	18.6 (36.9)	1.0 (0.7)
Narrow Passage (Difficult)	28.4 (26.0)	49.2 (66.7)	2.9 (1.8)
<b>Subjective Measures (Scale range)</b>	<b>Visual</b>	<b>Vibrotactile</b>	<b>Force Feedback</b>
Workload (0-20)	8.5 (3.7)	7.7 (3.0)	6.2 (2.5)
Spatial Orientation (1-7)	3.3 (1.6)	4.0 (1.5)	5.2 (1.2)
Collision Resolution (1-7)	3.6 (1.6)	3.9 (1.7)	2.8 (1.6)
Feedback Clarity (1-7)	4.7 (1.6)	3.4 (1.4)	5.7 (1.2)
System Usability (0-100)	72.6 (17.1)	68.6 (15.2)	82.2 (12)

*Standard deviations in parentheses*

## 5 Discussion

In the presented evaluation study, we compared visual, vibrotactile and force feedback for collisions in virtual environments with a generic VR assembly paradigm, including peg-in-hole and narrow passage tasks. Based on performance data, we found that the force feedback system with a light weight robot as input device and high resolution 6 DoF force feedback is superior in terms of precision compared to the vibrotactile and visual feedback systems. In all tasks, the applied forces were lowest when working with the haptic interface. Obviously, the high degree of haptic realism together with the fact that users are prevented from penetrating the virtual structures by force feedback contributed to higher manipulative performance. Yet, results also provided evidence for a potential trade-off between qualitative (movement precision) and quantitative (execution time) performance dimensions when using force feedback. In case of complex or multiple collisions with minimal clearances like it was the case in the difficult peg-in-hole task, users needed significantly more time to complete the task. Participants had problems when the virtual object was locked in the hole and the haptic interface could not be moved freely (like it would be the case in a real assembly task).

One potential drawback of many force feedback systems is that users are impeded from reaching a desired position quickly, since the haptic input device has to be moved (Aleotti et al., 2005). Yet, the overall pattern of completion times did not provide any evidence for this. In post-experimental interviews, some VR experts emphasized that this might even be an advantage, because the required input forces created



an illusion of object inertia and the interface also served the function of an arm rest, cf. [22]. Altogether, the overall usability was rated best. Subjective data also revealed that mental workload was rated substantially lower when working with the FF compared to the other systems. Haptic feedback was easy to interpret and individuals were able to react quickly, developed a high degree of spatial orientation and rarely had problems resolving collisions, i.e., the system was most supportive to build a mental picture of the virtual scene.

Substituting force feedback with vibrotactile information was cognitively more demanding due to feedback ambiguity. Seemingly, vibrotactile information mapping and density sometimes was confusing, leading to increased mental workload and loss of spatial orientation. Accordingly, performance data in the more difficult trials revealed that users took the “quick and dirty” approach, i.e., completion times were lowest as well as movement precision. Vibrotactile devices could be a reasonable alternative if high resolution of haptic information is not critical and/or to complement lacking visual information.

Compared to vibrotactile information, visual collision feedback was perceived as less ambiguous. Yet, completion times were higher during trials with unobstructed view, presumably because subjects had to process visual information of the virtual scene and collision visualization simultaneously. Therefore, users focussed on matching the guided and the feedback object instead of concentrating on trajectory planning.

We compared three different perception channels to provide haptic information in a virtual environment to the user: tactile, kinaesthetic (force feedback), and visual information. In future studies, bimodal feedback and combinations of interfaces should be explored.

## References

1. Burdea, G., Coiffet, P.: *Virtual reality technology*, 2nd edn. Wiley (2003)
2. Jayaram, S., Connacher, H., Lyons, K.: Virtual assembly using virtual reality techniques. *Comput. Aided Design* 29(8), 575–584 (1997)
3. Zorriassatine, F., Wykes, R., Parkin, R., Gindy, N.: A survey of virtual prototyping techniques for mechanical product development. *Proceedings Inst. Mech. Engineers. Part B: J. Eng. Manuf.* 217(4), 513–530 (2003)
4. Sung, R., Corney, J., Clark, D.: Automatic assembly feature recognition and disassembly sequence generation. *J. Comput. Inf. Sci. Eng.* 1(4), 291–299 (2001)
5. Seth, A., Vance, J., Oliver, J.: Virtual reality for assembly methods prototyping: A review. *Virtual Reality* 15(1), 5–20 (2011)
6. Burdea, G.: Haptics issues in virtual environments. In: *Computer Graphics International*, Geneva, Switzerland (2000)
7. Bergamasco, M.: The GLAD-IN-ART project. In: *Proceedings IMAGINA 1992*, II, pp. 7–14 (1992)
8. Lécuyer, A., Megard, C., Burkhardt, J.-M., Lim, T., Coquillart, S., Coiffet, P., et al.: The effect of haptic, visual and auditory feedback on an insertion task on a 2-screen workbench. In: *Proceedings of the Immersive Projection Technology (IPT) Symposium* (2002)

9. Cheng, L.-T., Kazman, R., Robinson, J.: Vibrotactile Feedback in Delicate Virtual Reality Operations. In: *ACM Multimedia*, pp. 243–251 (1996)
10. Zachmann, G., Gomes de Sa, A., Jakob, U.: Virtual Reality as a Tool for Verification of Assembly and Maintenance Processes. *Computers and Graphics* 23(3), 389–403 (1999)
11. Petzold, B., Zaeh, M.F., Färber, B., Deml, B., et al.: A study on visual, auditory and haptic feedback for assembly tasks. *Presence: Teleoperators and Virtual Environments* 13(1), 16–21 (2004)
12. Aleotti, J., Caselli, S., Reggiani, M.: Evaluation of Virtual Fixtures for a Robot Programming by Demonstration Interface. *IEEE Transactions on System Man, and Cybernetics, Part A: Systems and Humans* 35(4) (2005)
13. Lindeman, R., Templeman, J., Sibert, J., Cutler, J.: Handling of Virtual Contact in Immersive Virtual Environments: Beyond Visuals. *Virtual Reality* 6(3), 130–139 (2002)
14. Massimino, M.: Sensory Substitution for Force Feedback in Space Teleoperation. MIT Ph.D. Thesis, Department of Mechanical Engineering (1992)
15. Kontarinis, D., Howe, R.: Tactile Display of High-Frequency Information in Teleoperation and Virtual Environments. *Presence*, 387–402 (1995)
16. Bloomfield, A., Badler, N.: Virtual Training via Vibrotactile Arrays. *Presence: Teleoperators and Virtual Environments* 17(2), 103–120 (2008)
17. Burdea, G.: Haptic feedback for virtual reality. In: *Virtual Reality and Prototyping Workshop*, Laval, France (1999)
18. Volkov, S., Vance, J.: Effectiveness of Haptic Sensation for the Evaluation of Virtual Prototypes. In: *ASME Design Engineering Technical Conference*, Pittsburgh, PA (2001)
19. Schätzle, S., Ende, T., Wuesthoff, T., Preusche, C.: VibroTac: An ergonomic and versatile usable vibrotactile feedback device. In: *IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man)*, Viareggio, Italy (2010)
20. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*. North-Holland Press, Amsterdam (1988)
21. Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) *Usability Evaluation in Industry*. Taylor and Francis, London (1996)
22. Weber, B., Hellings, A., Tobergte, A., Lohmann, M.: Human Performance and Workload Evaluation of Input Modalities for Telesurgery. In: *Proceedings of the German Society of Ergonomics (GfA) Spring Congress*. GfA-Press, Dortmund (2013)

## **Part III**

# **Human-Robot Interaction in Virtual Environments**

# What Will You Do Next? A Cognitive Model for Understanding Others' Intentions Based on Shared Representations

Haris Dindo and Antonio Chella

University of Palermo, RoboticsLab, DICGIM, 90138 Palermo (PA), Italy  
{haris.dindo,antonio.chella}@unipa.it  
<http://roboticslab.dicgim.unipa.it>

**Abstract.** Goal-directed action selection is the problem of what to do next in order to progress towards goal achievement. This problem is computationally more complex in case of joint action settings where two or more agents coordinate their actions in space and time to bring about a common goal: actions performed by one agent influence the action possibilities of the other agents, and ultimately the goal achievement. While humans apparently effortlessly engage in complex joint actions, a number of questions remain to be solved to achieve similar performances in artificial agents: How agents represent and understand actions being performed by others? How this understanding influences the choice of agent's own future actions? How is the interaction process biased by prior information about the task? What is the role of more abstract cues such as others' beliefs or intentions?

In the last few years, researchers in computational neuroscience have begun investigating how control-theoretic models of individual motor control can be extended to explain various complex social phenomena, including action and intention understanding, imitation and joint action. The two cornerstones of control-theoretic models of motor control are the goal-directed nature of action and a widespread use of internal modeling. Indeed, when the control-theoretic view is applied to the realm of social interactions, it is assumed that *inverse* and *forward* internal models used in individual action planning and control are re-enacted in *simulation* in order to understand others' actions and to infer their intentions. This *motor simulation* view of social cognition has been adopted to explain a number of advanced *mindreading* abilities such as action, intention, and belief recognition, often in contrast with more classical cognitive theories - derived from rationality principles and conceptual theories of others' minds - that emphasize the dichotomy between action and perception.

Here we embrace the idea that implementing mindreading abilities is a necessary step towards a more natural collaboration between humans and robots in joint tasks. To efficiently collaborate, agents need to continuously estimate their teammates' proximal goals and distal intentions in order to choose what to do next. We present a probabilistic hierarchical architecture for joint action which takes inspiration from the idea of motor simulation above. The architecture models the casual relations between observables (e.g., observed movements) and their hidden causes

(e.g., action goals, intentions and beliefs) at two deeply intertwined levels: at the lowest level the same circuitry used to execute my own actions is re-enacted in simulation to infer and predict (proximal) actions performed by my interaction partner, while the highest level encodes more abstract task representations which govern each agent's observable behavior. Here we assume that the decision of what to do next can be taken by knowing 1) what the current task is and 2) what my teammate is currently doing. While these could be inferred via a costly (and inaccurate) process of inverting the generative model above, given the observed data, we will show how our organization facilitates such an inferential process by allowing agents to *share* a subset of hidden variables alleviating the need of complex inferential processes, such as explicit task allocation, or sophisticated communication strategies.

**Keywords:** joint action, motor simulation, shared representations, human-robot collaboration.

## 1 Introduction

Consider two agents (being human or artificial) collaborating on a joint task (e.g. building something together). How do they coordinate their actions without previous agreements or conventions? How do they adapt their actions during task execution? How do they achieve their goals? What are the computational mechanisms behind social interactions and joint action?

Here we argue that collaborative tasks (and social interaction problems, in general) require that interacting agents solve complex *mindreading* problems such as action and intention understanding, in parallel with motion planning and control. Indeed, recent research in social neuroscience has revealed that understanding the intentions of co-actors and predicting their next actions are fundamental for successful social interactions (cooperative or competitive) and joint actions [1,2]. In joint task such as building something together or running a dialogue, predictive mechanisms help the real-time coordination of one's own and the co-actor's actions and contribute to the success of the joint goal [3,4].

In the last years, there has been an increasing interest in joint action in the fields artificial intelligence and robotics (for a survey of related works see [5,6,7,8]) with the goal to make human-robot (or human-machine) collaboration increasingly more natural. To this aim, early researchers in AI have recognized the necessity to explicitly address the role of abstract social cues, such as intentions and beliefs, to efficiently handle teamwork problems [9]. Since then, various approaches have been built by adapting tools from symbolic reasoning [10], probabilistic decision processes [11], game theory [12] or by adopting a holistic approach based on cognitive architectures [13].

However, action understanding and prediction are hard (and often under-constrained) computational problems, and it is still unclear how humans solve them in real time while at the same time planning their complementary (or competitive) actions. It has been argued that action and intention recognition are

facilitated in joint action (but also more in general in social set-ups) because co-actors tend to automatically align (at multiple levels, of behavior and of cognitive representations), imitate each other, and share representations; in turn, this facilitates prediction, understanding, and ultimately coordination [14,15,16,17].

Here we present a computational (Bayesian) account for joint action, in which two or more agents act together so to realize a common goal. Inspired by ideas from computational neuroscience, our model describes joint action as a hierarchical phenomenon: (1) at the higher level, agents have to understand actions executed by other agents and their associated goals, and select actions that are complementary to those of the other agents or at least do not conflict with them; (2) at the lower level, agents have to coordinate their actions in real time and this requires a precise estimation of the timing and trajectories that is not necessary at the high level. Our model postulates that (shared) cognitive variables, such as beliefs and intentions, govern the activity of the motor system involved both in executing own actions and perceiving and understanding that of others via a *motor simulation* process. The following section provides a scientific background of our approach.

## 1.1 Background

Recognizing *what* another agent is doing and *why* (i.e., its distal intention) is extremely useful in social scenarios, both cooperative and competitive. Humans (and other animals adapted to social scenarios) are equipped with mechanisms for predicting and recognizing actions executed by others, inferring their underlying intentions, and planning actions that are complementary to them. An important constituent of the social mind of humans and monkeys is a neural mechanism for motor resonance, or the mapping observed actions into one's own motor repertoire: the *mirror system* [14]. This mechanism is part of a wide brain network that gives access to the cognitive variables (e.g., action goals and prior intentions) of another individual and permits to reconstruct the generative process that it uses to select the observed movements [16].

In this vein, it has been suggested that control-theoretic models of individual motor control can be extended to explain complex phenomena in social cognition [18,19]. The two cornerstones of control-theoretic models of motor control are the goal-directed nature of action, and the widespread use of internal modeling [20]. Indeed, when the control-theoretic view is applied to the realm of social interactions, the core scientific hypothesis is that these can be expressed through the overt and covert activity of predictive (i.e. forward) and prescriptive (i.e. inverse) internal models used in individual action planning and control [21]. In other words, an observing or interacting agent puts itself in others' shoes and elicits its own goal-directed representations in simulation to provide an embodied explanation of others' behavior. Apparently unrelated phenomena such as motor control [22], affordance recognition [23], imitation learning [24], action understanding [25], and joint action [26] - just to name a few - can efficiently and parsimoniously be explained by the process of internal re-enactment of one's own motor apparatus: a forward model can be used as simulator of the

consequences of an action, and when paired with an inverse model a degree of discrepancy between what I observe and what I do (or just “imagine” of doing) can be produced affording better understanding of their underlying goal [21,27]. These mechanisms of *motor simulation* could act in concert with other cognitive processes such as those regulating social attention, as well as with more demanding and deliberate ones, such as those that provide a full “theory of mind” [28].

Action understanding can be related to the estimation of the (most likely) current action another agent is performing, while deeper forms of mindreading can be associated to the inference of its intentions and beliefs. According to motor theories of cognition, the same architecture used for action planning and execution can be reused for understanding actions performed by others, and their underlying intentions. In addition to these high-level problems, the low level details of action specification, prediction and adaptation are solved on-line once a motor primitive is selected. However, low-level processes can influence the choice of cognitive variables, too. Indeed, the interplay between the two levels is bidirectional: the temporal unfolding of high-level constructs biases the action recognition process which, in turn, provides necessary information to monitor the execution of the joint task itself.

From a computational viewpoint, our model of mindreading implements the idea of competition between coupled inverse and forward models [27,21], but uses approximate Bayesian inference for solving the problem. A different proposal is that of [11], in which action understanding is realized through “inverse planning” methods, and for this reason is more closely related to the idea of teleological reasoning [29] than to the idea of motor simulation that we have put forward. Our model of joint action is related to the probabilistic model of [30] in that it includes a hierarchy of representations, but it also emphasizes the formation of shared representations and their role in guiding inferential processes. Finally, our analysis is related to other initiatives that investigated the neurocognitive mechanisms that make joint action so easy [31].

It emerges from our discussion that actions of an agent engaged in joint activities are governed by a continuous process of (joint) goal pursuing and adaptation to (1) the environment with its contextual constraints, and (2) the physical and interpersonal constraints offered by the actions of the co-actor and its abilities. The interplay of deliberate processes, which act on longer time scales, and faster processes of adaptation to the environment and the others, points to hierarchical models of action organization, with motor elements that belong to multiple levels of hierarchy (and give rise to processes that have different duration in time).

## 1.2 Are Shared Representations the Key for Successful Joint Actions?

Even if we assume the aforementioned hierarchical organization of action, it is currently unknown how the brain solves high- and low-level problems of joint action in real-time, given that their complexity is high even in simple scenarios [11].

We propose that co-actors do not solve interaction problems in isolation, but rather *with* the others (as well as with the environment): co-actors align their cognitive variables (beliefs, intentions and actions) and form *shared representations* (SR). We argue that what is shared during an interaction are the same representations for action (beliefs, intentions and actions) as used in individualistic action selection, performance and monitoring. For this model to work, it is not necessary that co-actors maintain separated representations for their own and another's actions, additional "we-representations", or meta-representations of what is shared. Rather, we call "shared" the subset of action representations that become aligned during interaction, being the co-actors aware of it, or not.

A first advantage of SRs is that the same cognitive variables can be used for action execution and prediction of another's actions (as well as for monitoring of the joint goal). Second, by sharing representations, an agent can help the other to understand and predict its own actions, and to select the next action to take; although this would not be optimal from an individualistic viewpoint, it can become so if the two agents are pursuing a joint action<sup>1</sup>.

From a computational viewpoint, shared representations help solving interaction problems in that they afford an *interactive strategy* for coordination that makes action selection and understanding easier. Put in simple terms, each agent involved in the joint action can:

1. Use motor simulation to infer what the other agent is doing (i.e., its actions) and why (up in the hierarchy of actions and intentions);
2. Infer which belief (and thus the associated sequence of intentions and actions) is the most likely one given the observed action, and 'align' its own belief;
3. Predict what is likely to happen next by using its own (chain of) intention and action representations, and in doing so, recognize affordances made possible (now or in the future) by the ongoing actions of the other agent;
4. Select complimentary (or successive) action by simply inferring what comes next in one's own intention and action representations (e.g., if I recognize that you are executing a certain action, I can start executing the next one in the sequence leading to the common goal);
5. While executing, lower level details are solved by other mechanisms of coordination and synchronization of action (e.g automatic entrainment, feedback, and motor simulation); in turn, as these mechanisms influence the choice of motor primitives, they have a bottom-up effect on the choice of cognitive variables;
6. When the confidence on the alignment of the joint goal is high - or when the details regarding the execution of the other agent are not essential - parts of this process can be skipped; for instance, in many circumstances co-actors can simply monitor the joint goal and use motor simulation only if an error is detected.

---

<sup>1</sup> An additional benefit of using shared representations is that, if each agent is confident that the other will facilitate it, for instance by signaling important events at the right time, then they can skip many costly mindreading and predictive processes.



We briefly mention that shared representations can be formed automatically or intentionally [32]. While in this paper we study automatic formation of shared representations, it is worth mentioning the role of *intentional* strategies that aim at influencing another’s cognitive variables so as to align them to one’s own. For instance, explicit communicative strategies such as the use of language, gesture, and deictics have the goal of forming or modifying shared representations. However, in [??] we focus on another - less studied - form of sensorimotor communication called *signaling*. Pushing a jointly-lifted table in a specific direction, over-articulating in noisy environment, and over-emphasizing vowels in child-directed speech are all examples of signaling. In all these examples, humans intentionally modify their action kinematics to make their goals easier to recognize. Thus, signaling acts in concert with automatic mechanisms of resonance, prediction, and imitation, especially when the context makes actions and intentions ambiguous and difficult to read. An in-depth discussion of how signaling helps joint interactions is out of the scope of the present paper (an interested reader can consult [26]).

Irrespective of how a shared representations are established, the common ground can be used as a coordination tool between two or more agents, like a blackboard in which two agents can read and write, which facilitates prediction of another’s behavior by drastically reducing uncertainty, and implicitly favors the unfolding of interactive sequences of behaviors in the two agents. It emerges from our analysis that the use of shared representations changes the nature of the (high level) interaction problem from the understanding and coordination with another’s actions to the active guidance of its beliefs, expectations and decisions. An agent can solve the problems of “what should I do next?” and “what will you do next?” by first inferring “what is the joint task?” and then using this information to solve the former problems. The next section provides a computational account of this process.

## 2 A Probabilistic Model of Joint Action

Social interaction in real world scenarios is an inherently stochastic process: perception and execution of motor acts are corrupted by noise and subject to failure, while planning of one’s own acts is subordinated to the recognition of others’ intentions and beliefs which are not directly observable. Furthermore, processes involved are tightly coupled (e.g. recognizing your goal-directed actions helps me updating my belief of the shared task being executed and predicting and anticipating your next steps).

We adopt the formalism of probabilistic graphical models embedding the idea of two levels of processing: at the lowest level the same circuitry used to execute my own actions are used to infer and predict the actions performed by my interaction partner via *motor simulation*, while at the highest level the two agents share action representations relative to the goals and tasks to be performed. The prior assumptions and beliefs about the joint task bias the action recognition process, while the specific motor acts confirms or disconfirms our current beliefs.

It is worth noting that two processes operate on different time scales: while the lowest level operates in real-time, providing an updated recognition of others' actions, the highest level is involved with less frequent transitions and it depends on the successful outcome of the lower levels.

In the next sections we will present our computational model focusing separately on the high- and low-level processes represented as Dynamic Bayesian Networks (DBNs). DBNs are Bayesian networks representing temporal probability models in which directed arrows depict assumptions of conditional (in)dependence between variables [33]. The general DBN model is defined by a set of  $N$  random variables  $\mathbf{Y} = \{Y^{(1)}, Y^{(2)}, \dots, Y^{(N)}\}$  and a pair  $\{BN^p, BN^t\}$  where  $BN^p$  represents the prior  $P(\mathbf{Y}_1)$  and  $BN^t$  is a two-slice temporal Bayesian network which defines

$$P(\mathbf{Y}_t | \mathbf{Y}_{t-1}) = \prod_{i=1}^N P(Y_t^i | Pa(Y_t^i)) \tag{1}$$

where  $Y_t^i$  is the  $i$ -th node at time  $t$  and  $Pa(Y_t^i)$  are the parents of  $Y_t^i$  in the graph (being in the same or previous time-slice). Usually, the variables are divided into hidden *state* variables,  $\mathcal{X}$ , and *observations*,  $\mathcal{Z}^2$ . From the computational point of view, the task of an *inference* process is to estimate the posterior joint distribution of hidden state variables at time  $t$ , given the set of observed variables so far<sup>3</sup>. By marginalizing the posterior distribution it is possible to answer questions about particular variables in the network (e.g. what is the probability that a particular motor act has been executed at time  $t$ ?). Next two sections provide an overview of our architecture for joint action (for a detailed description of various processes, and for an analysis of the experimental results, please consult [25,26]).

### 2.1 Low-Level Model

The low-level model implements a *motor simulation* process that guides perceptual processing and provides action recognition capabilities. In motor simulation, it is the reenactment of one's own internal models, both inverse and forward, used for interaction that provides an understanding of what others are doing.

The entire process of action understanding can be cast into a Dynamic Bayesian Network (DBN) shown in Figure 1(a). As usual, shaded nodes represent observed variables while others are hidden and need to be estimated through the process of probabilistic inference. The model embeds the idea of motor simulation by including a probabilistic representation of forward and inverse models activation. In our representation, the process of action understanding is influenced by the following factors expressed as stochastic variables in the model (fig. 1b):

1. *MP*: index of the agent's own repertoire of goal-directed motor primitives; each motor primitive directly influences the activation of related forward and inverse models;

---

<sup>2</sup> By convention the observed variables are represented as shaded nodes in the network.

<sup>3</sup> This process is also known as filtering.

2.  $u$ : continuous control variable (e.g. forces, velocities, ...);
3.  $x$ : state (e.g. the position of the demonstrator’s end-effector in an allocentric reference frame);
4.  $z$ : observation, a perceptual measurement related to the state (e.g. the perceived position of the demonstrator’s end-effector on the retina).

Figure 1c shows the conditional distributions which arise in the model. The semantics of the stochastic variables, and the concrete instantiation of the conditional distributions depends on the experimental setting. Suppose we can extract the noisy measurements of the true state of the demonstrator,  $z_t$ , through some predefined perceptual process described probabilistically by the observation model  $p(z_t|x_t)$ . Motor primitive index variable,  $MP$ , is associated with a paired inverse-forward model, and it implicitly encodes the demonstrator’s goal (in terms of the perceiver’s one). The initial choice of which internal models to activate is biased by the prior probabilities (here set by the high-level network). Each paired internal model  $MP_t$  is responsible of both generating a motor control  $u_t$ , given the (hidden) state  $x_{t-1}$  (inverse model), and of predicting the next (hidden) state  $x_t$ , given the motor control  $u_t$  and the previous state  $x_{t-1}$  (forward model).

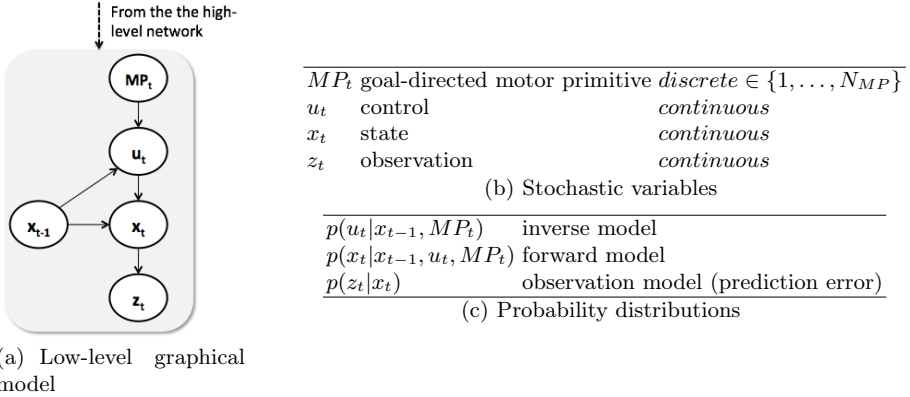
Given that in our model each goal-directed action is encoded as a coupled forward/inverse model, to predict and understand the actions performed by others it is sufficient to compute the posterior distribution over possible forward-inverse action pairs given all the observations so far,  $p(MP_t|z_{1:t})$ . This distribution can be obtained by marginalizing the full conditional posterior (i.e. belief) over all hidden variables in the model. Let us denote with  $\mathcal{X}_t$  the set of hidden variables at time  $t$ , and with  $\mathcal{Z}_t$  the set of observed variables at the same time step, the full conditional posterior can be obtained by the well-known recursive Bayesian inference schema [33]:

$$p(\mathcal{X}_t|\mathcal{Z}_{1:t}) = \eta p(\mathcal{Z}_t|\mathcal{X}_t) \cdot \int p(\mathcal{X}_t|\mathcal{X}_{t-1}) \cdot p(\mathcal{X}_{t-1}|\mathcal{Z}_{1:t-1}) d\mathcal{X}_{t-1} \quad (2)$$

where  $p(\mathcal{X}_t|\mathcal{X}_{t-1})$  and  $p(\mathcal{Z}_t|\mathcal{X}_t)$  are called prediction and observation models, respectively.

However, in order to compute the most likely observed action, the recursive propagation of the posterior density  $p(\mathcal{X}_t|\mathcal{Z}_{1:t})$  in equation 2 is only a theoretical possibility, and in general it cannot be determined analytically. By casting the problem of action prediction and understanding in a Bayesian framework permits to adopt efficient techniques for *approximate* probabilistic inference under the constraint of limited resources. We adopt *particle filters*, a Monte Carlo technique for sequential simulation [34]. The key idea of particle filters is to represent the required posterior density function by a set of random samples with associated weights and to compute probabilistic estimates of interested quantities based on these samples and weights. Each random sample is therefore a weighted hypothesis of an internal model activation in the action prediction task, where the weight of each particle is computed according to the divergence between the predicted state of the internal model the particle belongs to and the

observed state; intuitively, severe discrepancies between predictions produced by coupled internal models and observed percepts will lead to assigning low weights to internal models less involved in explaining the current action observation. Our approach permits to solve the problem of intention recognition in real-time under the assumption that what I am observing can be adequately explained through my own internal models. The particle filter schema allows to use a multitude of internal models, for various skills and contexts, and to focus only on those able to accurately explain the current observations [25].



**Fig. 1.** Graphical model (DBN) for action understanding based on coupled forward-inverse models; Adapted from [25]

## 2.2 High-Level Model

During observation of actions executed by others, motor simulation provides information that can be used to filter perceptual processing by allocating more resources (i.e., more particles in the particle filtering algorithm) to the most likely observations. This process achieves two objectives at the same time: first, it helps perceptual processing (like in Kalman filtering), and second, it permits to recognize the observed actions at the goal level by mapping them into the perceiver’s repertoire of internal models.

However, in order to initialize the low-level portion of the network, we need to set the prior probability distribution over the goal-directed internal model pairs. In a joint task this distribution should be estimated by a higher-order process connected with the more abstract task representation. Some motor acts, viewed as paired forward-inverse models, are more probable at a point in time during the execution of a particular joint task. Therefore, the high-level portion of our computational model should bias the action recognition process, while at the same time providing a parsimonious way to encode the shared representations. Additionally, the interplay between the low- and high-level portions of our network shall not be unidirectional: the recognition of others’ motor acts helps also

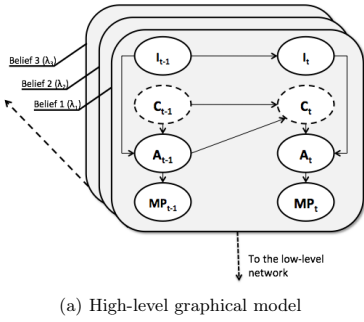
monitoring the joint act itself by revising hypotheses on the distal goal of the task in a similar vein as done in the low-level network. Here, recognized motor primitives act as observations for an abstract probabilistic representation of joint tasks and fitness agents' current belief. In addition, the high-level model provides a parsimonious way to encode shared representations as explained below.

In our computational model, a joint action is influenced by three main factors: intentions, contextual information (representing the observable state of the world and its affordances) and possible actions each actor can perform given the context and intentions. The temporal evolution of these factors can be represented once again by using the formalism of probabilistic graphical models (DBN). However, each joint task requires a different motor plan, and its representation should account for possible failures in the execution. For this reason, the high level portion of our computational model includes a battery of DBNs, each one representing a possible evolution of the joint task over time (figure 2(a)). A full DBN corresponds to a belief, which intuitively encodes knowledge of "what is the task we are performing?". The stochastic variables and conditional distributions of the high-level DBN are described in figure 2(b-c).

As an example, suppose two agents (e.g. a human and a robot) have to jointly build one out of several types of towers ( $\lambda_1, \lambda_2, \dots \lambda_n$ ) given a set of available red and blue blocks. Each high-level network (figure 2(a)) represents a particular type of tower and can be seen as implicitly encoding the beliefs each actor has regarding the execution of the task. For instance, the tower can be made of blocks having the same color (e.g. red or blue), or of two interleaved colors (e.g. red-blue-red-blue-...). The prior probability,  $p(\lambda)$  reflects the knowledge of which tower is more probable. The variable  $I_t$  models the intention to pick and place a block of a particular color onto the tower, while the contextual variable  $C_t$  could model the availability of red and blue blocks. The action variable  $A_t$  represents the action of manipulating a particular object in the world, and it directly influences the activation of motor primitives ( $MP_t$ ) used to efficiently execute the action. Motor primitives represent the observed variable and they are estimated by the low-level portion of our network at every step (see 2.1). Once an action is executed, the network models the transition to the next intention and next context through the corresponding transition probabilities.

We assume that the same set of models is shared across the two joint actors. However, their probabilistic parameters (prior, transition and observation probabilities) can be different according to individual actor's knowledge and expertise. The goal of the actors is to align their beliefs. From the probabilistic standpoint the machinery involved differs if the actor has to perform an action or if it has to recognize the action performed by another actor and update its belief. However, both computational problems have at its core the process of estimating the *likelihood* of each model given the observations.

If we denote the prior probability of a model as  $P(\lambda)$ , the goal is to compute the probability of the model given the set of observations so far (e.g. the likeli-



$I_t$	intention	<i>discrete</i> $\in \{1, \dots, N_I\}$
$C_t$	context	<i>discrete</i> $\in \{1, \dots, N_C\}$
$A_t$	goal-directed action	<i>discrete</i> $\in \{1, \dots, N_A\}$
$MP_t$	goal-directed motor primitive	<i>discrete</i> $\in \{0, \dots, N_{MP}\}$
$\lambda$	belief	<i>discrete</i> $\in \{1, \dots, N_\lambda\}$

(b) Stochastic variables

$p(I_t I_{t-1})$	intentional dynamics
$p(C_t C_{t-1}, A_{t-1})$	contextual dynamics
$p(A_t I_t)$	action induction
$p(U_t A_t)$	utility function
$p(MP_t A_t)$	motor primitive induction

(c) Probability distributions

**Fig. 2.** High-level battery of Dynamic Bayesian Networks (DBN) for joint-action. Every network in the battery is a probabilistic representation of the shared task. Adapted from [26].

hood):  $P(\lambda_i|MP_{1:t})^4$ . The most plausible model is the one that maximizes the posterior probability of the model:

$$\operatorname{argmax}_{\lambda_i} P(\lambda_i|MP_{1:t})P(\lambda_i), \forall i \in \{1, \dots, N_\lambda\} \quad (3)$$

The likelihood is used in both action recognition and selection. In action recognition, it is used to initialize the process of motor simulation; in action selection, it is used to choose the best action to perform so that it does not lower the current likelihood. The presence of shared representations permits to describe the process in an unconventional way. Specifically, both agents use the same high-level network, in which observed and executed intentions and actions are treated on a pair, independent on who executes them. Note that the same formulation can be used to model tasks in which two agents act synchronously, such as for instance when they lift together a block, and turn-based tasks, in which one agent acts at times  $t, t+2, t+4, \dots$  and another agent acts at times  $t+1, t+3, t+5, \dots$

The first part of the inference is the same for action observation and action selection: at each turn agents compute the likelihood of all the available models given all the observations so far (rather recognized or performed motor primitives,  $MP$ ), and the belief with the highest likelihood is treated as the goal state. Action observation is then implemented as a filtering process; first, the intention  $I_{t+1}$  belonging to the current belief is predicted, which is then used to bias the recognition of  $MP$  by accordingly setting the prior probabilities needed to trigger the low-level network activation. For instance, if the system believes that the task is to build a tower made of six red blocks, it predicts that the next intention ( $I_{t+1}$ ) will be to place a red block, and then it uses this information to bias the perception of actions executed by the other agent (i.e., the estimation of  $MP_{t+1}$ ). In turn, the lower level affects high-level goal selection, as prediction errors drive belief revision (this is typical of hierarchical generative

<sup>4</sup> Likelihood computation in this network can be performed exactly by the forward-backward algorithm or approximately by the abovementioned particle filters.

models [35,36]): the recognized action is treated as an observation for the high-level network, and it is used by the observing agent to revise its current belief and eventually to align its shared representation to that of the other actor by computing the current likelihood (cf. equation 3).

Action selection is different from action observation in that MP cannot be observed (in fact, it has to be produced). Still, the process is conceptually the same: first, the intention  $I_{t+1}$  belonging to the belief with the highest likelihood is predicted; then, the most probable  $MP$  is selected for execution. For instance, if the system believes that the task is to build a tower made of six red blocks, it first predicts the most probable next intention ( $I_{t+1}$ ) compatible with this belief (i.e., the intention to place a red block), then it generates an associated action (i.e., taking a specific red block), and finally an associated MP (i.e., the motor process for grasping the selected block).

### 3 Conclusions

Joint actions between humans and artificial agents are notoriously difficult to implement and the issue of what kind of cognitive processing is required in cooperation, coordination, and joint action is still debated. We postulate that joint actions (and social interactions in general) are heavily guided by abstract cognitive variables, such as goals, intentions and beliefs, and that the interaction itself is facilitated if interacting agents could have access to such variables. We present a computational account that allows agents to automatically align their internal representations (i.e., inferring “what task are we pursuing?” and choosing the hypothesis with higher likelihood) and then using this information in a generative scheme to both (i) decide what to do next, and (ii) predict what the other agent will do next. To cope with uncertainty, our model is developed as a two-level dynamic Bayesian network, where the lowest level implements the process of motor simulation to understand and anticipate other agent’s (proximal) action intentions, while the highest level provides an abstract encoding of the task and the (distal) goals. The two levels are deeply intertwined: the temporal unfolding of high-level constructs biases the action recognition process which, in turn, provides necessary information to monitor the execution of the joint task itself. In a nutshell, our model exports the ideas from individual motor planning, control and monitoring to the realm of social interactions, by adopting the the motor view of social cognition augmented with more abstract cognitive constructs that guide the interaction.

Since any observable behavior can generally be explained by many underlying intentions and beliefs, in order to disambiguate them it is necessary to adopt costly inferential processes. Part of this cost can be alleviated by forming shared representations (SR) and using them as a coordination tool. Here we do not investigate the origin of shared representations; we see SR as a blackboard in which two agents can read and write and which facilitates prediction of another’s behavior by drastically reducing the uncertainty of mindreading inferential processes.

**Acknowledgements.** The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement #231453 (HUMANOBS).

## References

1. Sebanz, N., Bekkering, H., Knoblich, G.: Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10(2), 70–76 (2006)
2. Newman-Norlund, R.D., Noordzij, M.L., Meulenbroek, R.G., Bekkering, H.: Exploring the brain basis of joint action: Co-ordination of actions, goals and intentions. *Social Neuroscience* 2(1), 48–65 (2007)
3. Sebanz, N., Knoblich, G.: Prediction in joint action: What, when, and where. *Topics in Cognitive Science* 1, 353–367 (2009)
4. Pickering, M.J., Garrod, S.: An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* (forthcoming)
5. Fong, T., Thorpe, C., Baur, C.: Collaboration, dialogue, human-robot interaction. *Robotics Research*, 255–266 (2003)
6. Breazeal, C.: *Designing sociable robots*. The MIT Press (2004)
7. Sun, R.: *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge University Press (2005)
8. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction* 1(3), 203–275 (2007)
9. Cohen, P.R., Levesque, H.J.: Teamwork. *Nous*, 487–512 (1991)
10. Breazeal, C.: Social interactions in hri: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 34(2), 181–186 (2004)
11. Baker, C.L., Saxe, R., Tenenbaum, J.B.: Action understanding as inverse planning. *Cognition* 113(3), 329–349 (2009)
12. Yoshida, W., Dolan, R.J., Friston, K.J.: Game theory of mind. *PLoS Comput. Biol.* 4(12), e1000254+ (2008)
13. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. *Psychological Review* 111(4), 1036 (2004)
14. Rizzolatti, G., Craighero, L.: The mirror-neuron system. *Annual Review of Neuroscience* 27, 169–192 (2004)
15. Frith, C.D., Frith, U.: How we predict what other people are going to do. *Brain Research* 1079(1), 36–46 (2006)
16. Kilner, J.M., Friston, K.J., Frith, C.D.: Predictive coding: An account of the mirror neuron system. *Cognitive Processing* 8(3), 159–166 (2007)
17. Pezzulo, G., Candidi, M., Dindo, H., Barca, L.: Action simulation in the human brain: Twelve questions. *New Ideas in Psychology* (2013)
18. Grush, R.: The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences* 27(3), 377–396 (2004)
19. Gardenfors, P.: Mind-reading as control theory. *European Review* 15(2), 223–240 (2007)
20. Wolpert, D.M., Ghahramani, Z.: Computational motor control. In: Gazzaniga, M. (ed.) *The Cognitive Neurosciences III*, pp. 485–494. MIT Press (2004)
21. Wolpert, D.M., Doya, K., Kawato, M.: A unifying computational framework for motor control and social interaction. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 358(1431), 593–602 (2003)



22. Jordan, M.I., Wolpert, D.M.: Computational motor control. *The Cognitive Neurosciences* 601 (1999)
23. Fitzpatrick, P., Metta, G., Natale, L., Rao, S., Sandini, G.: Learning about objects through action-initial steps towards artificial cognition. In: *Proceedings of IEEE International Conference on Robotics and Automation, ICRA 2003*, vol. 3, pp. 3140–3145. IEEE (2003)
24. Dindo, H., Schillaci, G.: An Adaptive Probabilistic Approach to Goal-Level Imitation Learning. In: *Proc. of the 2010 IEEE/RSJ International Conference on Intelligent RObots and Systems (IROS)*, October 18-22, pp. 4452–4457 (2010), doi:10.1109/IROS.2010.5654440
25. Dindo, H., Zambuto, D., Pezzulo, G.: Motor simulation via coupled internal models using sequential monte carlo. In: *Proceedings of IJCAI 2011*, pp. 2113–2119 (2011)
26. Pezzulo, G., Dindo, H.: What should I do next? using shared representations to solve interaction problems. *Experimental Brain Research* 211(3), 613–630 (2011)
27. Demiris, Y., Khadhour, B.: Hierarchical attentive multiple models for execution and recognition (hammer). *Robotics and Autonomous Systems Journal* 54, 361–369 (2005)
28. Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H.: Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28(5), 675–691 (2005); discussion 691–735
29. Csibra, G., Gergely, G.: Obsessed with goals: Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica* 124, 60–78 (2007)
30. Cuijpers, R.H., van Schie, H.T., Koppen, M., Erhagen, W., Bekkering, H.: Goals and means in action observation: a computational approach. *Neural Netw.* 19(3), 311–322 (2006)
31. Vesper, C., Butterfill, S., Knoblich, G., Sebanz, N.: A minimal architecture for joint action. *Neural Networks* 23(8-9), 998–1003 (2010)
32. Knoblich, G., Sebanz, N.: Evolving intentions for social interaction: from entrainment to joint action. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 363(1499), 2021–2031 (2008)
33. Murphy, K.P.: *Machine Learning: A Probabilistic Perspective*. MIT Press (2012)
34. Doucet, A., Johansen, A.M.: A tutorial on particle filtering and smoothing: fifteen years later. In: *Handbook of Nonlinear Filtering*, pp. 656–704 (2009)
35. Friston, K.: Hierarchical models in the brain. *PLoS Computational Biology* 4(11), e1000211 (2008)
36. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2(1), 79–87 (1999)

# Toward Task-Based Mental Models of Human-Robot Teaming: A Bayesian Approach

Michael A. Goodrich and Daqing Yi

Brigham Young University, Provo, UT, 84602, USA  
mike@cs.byu.edu, daqing.yi@byu.edu

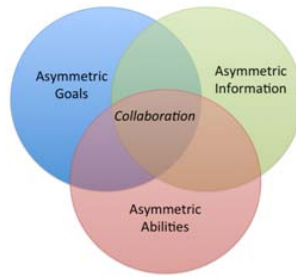
**Abstract.** We consider a set of team-based information tasks, meaning that the team's goals are to choose behaviors that provide or enhance information available to the team. These information tasks occur across a region of space and must be performed for a period of time. We present a Bayesian model for (a) how information flows in the world and (b) how information is altered in the world by the location and perceptions of both humans and robots. Building from this model, we specify the requirements for a robot's computational mental model of the task and the human teammate, including the need to understand where and how the human processes information in the world. The robot can use this mental model to select its behaviors to support the team objective, subject to a set of mission constraints.

## 1 Introduction

In complex, rapidly evolving team settings in which a robot fulfills a role, the robot needs sufficient autonomy to allow its human teammates to be free to direct their attention to a wider range of mission-relevant tasks that may or may not involve the robot. In contrast to many prior applications in which the robot was either teleoperated or managed under strictly supervisory control [1], recent advances in robot technologies and autonomy algorithms are making it feasible to consider creating teams in which a robot acts as a teammate rather than a tool [2].

In this team-centered approach, both humans and robots can take on roles that match their strengths. Properly designed, this can facilitate the performance of the entire team. This idea has already been applied to reform human-robot interaction in many areas, like object identification, collaborative tasks performance, etc. [3]. In this paper, we adopt the notion of collaboration, operationally defined as the process of utilizing shared resources (communication, space, time) in the presence of asymmetric goals, asymmetric information, and asymmetric abilities as illustrated in Fig. 1. The word collaboration suggests that there are both overlaps and differences between the goals, information, and abilities of the agents involved. Colloquially, collaboration can happen when everyone has something unique to offer and something unique to gain, but there is some benefit to each individual if activity is correlated.

In a human-robot team, the asymmetries on abilities and information mostly come from the natural difference on agents' sensors and actuators. Additionally, an agent may exhibit ability and information asymmetry in different states of interacting with



**Fig. 1.** Operational Elements of Collaboration

the environment, like location, lighting condition etc. Often, a team goal will be decomposed into subgoals in execution. The subgoals are usually assigned to agents in the team by organizing agents into specific roles with specific responsibilities, and this leads to goal asymmetries. In a collaboration framework, the interaction between agents not only focuses on common goals, but may also require providing support for others' goals. In a team search tasks, for example, the robot and the human might work together for target searching, while the robot might assist the human to deal with an emergency.

Collaboration is a form of teamwork that benefits from an explicit representation of shared intent. The theory of shared intent suggests that both the human and the robot need to have a mental model for the task to be performed and another mental model for how other team members will act [4]. The primary contribution of this paper is a framework for developing a task-based mental model from a human-robot collaboration perspective, including the ability to represent and reason about contributions of other team members to the mission and estimation of how other team members' actions affect performance.

## 2 Shared Mental Model

From studies of cognitive psychology, the concept of a shared mental model has been proposed as a hypothetical construct, which has been used to model and explain certain coordinated behaviors of teams. Shared mental models provide a framework of mutual awareness, which serves as the means by which an agent selects actions that are consistent and coordinated with those of its teammates. According to [5] [6], in order to perform collaboratively as a team, members of a team must have the following:

- **Teammate Model:** knowledge of teammates skills, abilities and tendencies.
- **Team Interaction Model:** knowledge of roles, responsibilities, information sources, communication channels and role interdependencies.
- **Team Task Model:** knowledge of procedures, equipment, situations, constraints.

These elements determine (a) how an agent makes decisions as a member of the team and (b) how diverse capabilities and means of interactions are managed within an organizational context. These concepts have been incorporated as important elements in

existing human-robot team designs [7]. From a robot's perspective, operating within the context of a human-robot team, the robot's shared mental model will help the robot predict information and resource requirements of its teammates. Importantly, a better understanding of task demands and how teammates will likely respond will enhance the robot's ability to support team-level adaptations to changes in the world.

Given a shared intent from the team, the robot is assigned or adopts tasks, either as an autonomous agent or as a collaborating teammate. What does the robot need to collaborate? We address two fundamental elements: (1) How should the robot model the task? (2) How should the robot model a human performing the task? We then illustrate how the concept of a shared mental model is applied within a search task by providing an example computational model that responds to these two questions.

### 3 Robot Wingman in a Search Task

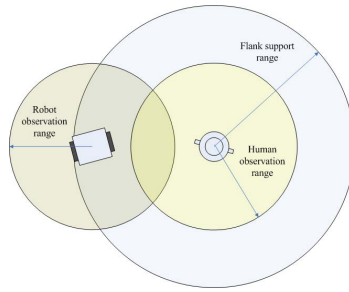
We introduce the shared mental model to a human-robot team search problem. In the problem, the search region is modeled with the belief of where the target objects are, and the search process works as constantly updating this belief by observations. Thus, teams of humans and robots manage a region of space subject to particular time or timing constraints [8].

From prior work in search theory, search efficiency is usually considered as one of the essential factors to a task success, and is therefore a central element of the team's model of a search task. There are several parameters to measure the efficiency in a search task [8], which are determined by the observation capability of a search agent. In this paper, we are interested in:

- **Sweep Width:** a measure of how wide an area a searcher can, on average, effectively cover. More specifically, it represents how well a sensor (e.g., the human eye) can detect specific objects as a function of distance from sensor to object.
- **Coverage:** a simple measure of how well a segment was covered by all of the searchers. Coverage is a ratio calculated by summing up the area that each searcher covered and dividing by the area of the search segment.
- **Probability of Detection:** a measure of the probability of success. Search managers need a way to determine the probability that a lost object would have been found if it was actually in the segment that was searched.

*Effective swept width* and *coverage* are determined by the sensor model of a search agent; the sensor model encodes the characteristics and capabilities of the agent's sensors. This model defines what the observation range of an agent is, and how the observation uncertainty might change with the distance of a target object. By contrast, the *probability of detection* shows the probability that an object would have been detected if in the area, which can be modeled as (a) an agent's prior belief that an object is in the search region and (b) the quality of the agent's observation. Since all agents are imperfect detectors, there exist differences in detection success and detection times among agents.

There is relevant prior work on applying these concepts to human-robot teams. [9] and [10] import robots into urban search and rescue so that human unreachable locations



**Fig. 2.** A Robot Wingman framework

can be explored, which greatly extends the coverage of search task execution. Integrating various types of sensors, like radars, laser rangefinders, ultrasonic sensors etc. [11] [12], greatly expands the sweep width of a search team. [13] and [14] propose a way to improve the probability of detection using information fusion across multiple agents. Modeling the team as a distributed information fusion process exploits the asymmetric perception capabilities of humans and robots to enhance the search efficiency of the team.

In our proposed human-robot search team, we assume that the human is better at strategy and decision making and the robot is better at raw data collection. This assumption forms the basis for the robot's model of its teammate. We propose the notion of a *robot wingman* to support a human in a collaborative search task, which is to have a robot that accompanies a human as he or she navigates through some space. Since a robot may be able to detect certain types of signals not perceivable by a human (e.g., radio signals or chemical gradients), it is possible for the wingman robot to extend the team's perception not only in space but also in the type of data perceivable by the team. As shown in Fig.2, as a flank support range that constrains where the wingman can move. The robot wingman is expected to stay in an area determined by the flank support range around the human, when the human is moving for the search task. Doing so guarantees that the robot rapidly respond to the human needs assistance, which maintains a reasonable distance for supporting communication and coordination. In a shared human-robot search problem, the robot's role not only includes staying within the flank support range, but also includes gathering information about the world around the team.

The organization of the team determines how information flows when executing the search task. Thus, the organization is an important element of the team interaction model, with information flow acting as the currency of interaction. Information flow shapes the process of fusing asymmetric information for collaboration. In the next section, we model the belief of the locations of the search objects by a shared task model. The information comes from the observations from both the human and the robot. Meanwhile, the robot predicts how the human will work and what the information collected by human is like, and this prediction is used to make a decision on how to run the search operations as in Fig.3.

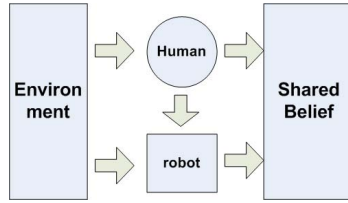


Fig. 3. Information Flow in a Wingman Human-Robot team

### 4 A Bayesian Approach

We present a Bayesian model for how information flows in the world and how information is altered in the world by the locations and perceptions of both humans and robots. Building from this model, we can specify the requirements for a computational mental model of the human teammate to understand where and how the human processes information in the world. The robot can then select its behaviors to support the team objective, subject to a set of mission constraints.

The world is represented as a discrete set of cells. For each cell, we wish to determine the probability that an object of interest is in a particular cell given a set of observations. Let  $S_i^t$  and  $O_i^t$  denote state and observation random variables that encodes whether an object of interest is in cell  $i$  at time  $t$ . Given a set of  $N$  cells, we will move or position the robot such that we gather a series of observations that provide information about all of the cells or some subset of those cells.

Since observations will be taken over time and since objects of interest can move over time, we formulate the problem as a sequential Bayes estimation problem. Given  $t$  sequential observations about cell  $i$ , our belief that an object of interest is in cell  $i$  at time  $t$  is given by the following:

$$bel^t(s_i) = P_{S_i^t | O_i^t, O_i^{t-1}, \dots, O_i^1}(s_i^t | o_i^t, o_i^{t-1} \dots o_i^1). \tag{1}$$

Equation (1) is the a posteriori estimate that an object of interest in cell  $i$  has been detected given all observations to that point position. Adopting the standard conditional independence assumptions of the Bayes filter [15], the sequential estimate becomes

$$bel^t(s_i) = \alpha P_{O_i^t | S_i^t}(o_i^t | s_i) \overline{bel}^t(s_i), \tag{2}$$

$$\overline{bel}^t(s_i) = \sum_j \sum_{s_j \in S} [P_{S_i^t | S_j^{t-1}}(s_i | s_j) bel^{t-1}(s_j)], \tag{3}$$

where  $\overline{bel}^t(s_i)$  is the predicted distribution of objects of interest,  $\alpha$  is the normalizing constant required by Bayes rule (equal to one divided by the prior predictive distribution),  $P_{O_i^t | S_i^t}(o_i | s_i)$  is the detection likelihood, and  $P_{S_i^t | S_j^{t-1}}(s_i | s_j)$  is the model for how objects move in the world.

In this paper,  $s = T$  or  $s = F$  indicate that the cell contains an object of interest or not. For each cell, we track the belief that an object of interest is in that cell as a

function of time. Given a prior belief about objects in the cell, we predict the probability that an object of interest will still be in that cell given (a) the presence or absence of an object in that cell on the previous time step, and (b) the presence or absence of objects in neighboring cells in the previous time step. Thus, Equation (3) includes a double summation, one for all cells in the world (the sum over  $j$ ) and the other over the presence or absence of objects in that cell.

The process of a search task can also be considered as information gathering. From (2) and (3), we can see that information from observation updates the belief of the search region, which results in uncertainty reduction. We select entropy, which is a commonly used criterion for measuring uncertainty [16], to quantify information collection. It is written as:

$$H(\text{bel}^t(s_i)) = - \sum_{s_i \in S} [\text{bel}^t(s_i) \log(\text{bel}^t(s_i))]. \quad (4)$$

## 5 Case Study

Consider a two-dimensional simplified representation of the world and adopt an occupancy grid representation of information in the world. We create a hexagonal tessellation of the world with the dimension of the hexagon determined by the perceptual capabilities of the human. The hexagonal tessellation is useful because it is one in which the distance from the center of one cell to any of its immediate neighbors is constant.

Before exploring the search region, we have no information on this area. We use the entropy of the shared belief to define the uncertainty in equation (4). More formally, we will assume that the prior probability that a cell is occupied by an object of interest is equal to 0.5, which means that the probabilities of the search object in the cell or not are equivalent.

### 5.1 Teammate Model

From the teammate model of human behavior, the wingman robot can predict how the human will move. This yields a sequence of cells that the human plans to traverse, which is denoted by  $Y = [y^1; y^2; \dots; y^D]$ . Each  $y^t$  corresponds to a physical location in the tessellation, so  $y^t = i$  means that the human was in cell  $i$  at time  $t$ .

We adopt a very simple model of agent perception, albeit one based in search theory. The model is that the likelihood of detecting an object of interest in cell  $i$  is certain if the human occupies that cell, is zero for cells outside a fixed radius of detection, and is constant for all cells within the radius of detection. Let  $N(i)$  denote the set of all cells that are within  $R$  units of cell  $i$ , in which  $R$  defines a radius,

$$N(i) = \{j : j \text{ is no further than } R \text{ cells from } i\}. \quad (5)$$

Let  $\lambda \in (0, 1)$  be the constant of detection for all cells within  $N(i)$ . Thus, an agent's probability of detection at position  $x^t$  is given by Equation (6).

$$P_{O_i^t | S_i^t}(F | T) = \begin{cases} 0 & \text{if } i = x^t \\ 1 - \lambda & \text{if } i \in N(x^t) \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

By definition,  $P_{O_i^t|S_i^t}(T | T) = 1 - P_{O_i^t|S_i^t}(F | T)$ . In (6), we assume a search agent can do perfect observation in the cell he is in. However, there exist distinctions on probabilities of detection in the neighbor cells, which come from the difference on agent perception capabilities. To differentiate the observation range, we use  $N^{human}(y^t)$  and  $N^{robot}(x^t)$  for the set of observed cells by human and robot at time  $t$ .

### 5.2 Team Interaction Model

The team interaction model uses the flank support range  $R_{flank}^{human}$  to determine the set of cells for the wingman robot motion. This is based on the human’s tolerance for how far the robot can wander before being out of position. We use (5) to translate  $R_{flank}^{human}$  into a set of feasible cells,  $N_{flank}^{human}(y^t)$ , in which  $y^t$  is the human’s position. Given a motion range of the wingman robot at a time step,  $R_{motion}^{robot}$ , we have

$$\forall y^t, x_t \in N_{flank}^{human}(y^t) \cap N_{motion}^{robot}(x^{t-1}) \tag{7}$$

to define the wingman robot motion dependence on the human motion.

In this paper, we assume that the teammate model provides enough information to estimate the human’s path via prediction, Equation (6) can be used to determine the posterior probability of likely target location after the human has moved. The posterior from the human is then used as the prior for the robot. In essence, this means that the shared belief about the world passes through two phases: a refinement that comes because the human has moved through the environment and a refinement that comes because the robot is going to move through the environment.

### 5.3 Team Task Model

When the robot plans to fulfill its role for the task model, it assumes that objects of interest do not move, appear, or disappear over time, but this will change in future work. Given this assumption, the prior estimate for the target object’s location at time  $t$  is equal to the posterior estimate for target object location at time  $t - 1$ . In future work, if the object of interest can move, then a predictive step is required and a full Bayes filter can be applied [15].

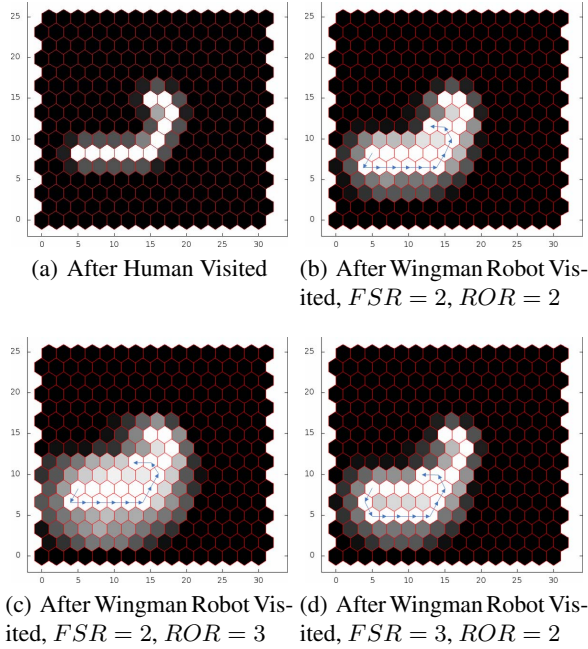
Since the human’s path has been obtained from the teammate model (the robot is supporting the human), our goal is to control the robot’s path to maximize the amount of information gathered by the human and robot combined. When the robot is at  $x^t$ , it will update the beliefs of all the neighbor cells defined by the radius of detection,  $R^{robot}$ . We denote the information gain at position  $x^t$  as

$$F(x^t) = \sum_{i \in N^{agent}(x^t) \cup x^t} [H(bel_i^{t-1}) - H(bel_i^t)]. \tag{8}$$

In (8),  $H(bel_i^t)$  denotes the entropy of the belief of cell  $i$  at time  $t$ , and  $F(x^t)$  shows the uncertainty reduction at time  $t$  from the all the observed cells.

In order to keep synchronization with human motion, a requirement imposed by the interaction and task model, we assume that the robot starts in the same location





**Fig. 4.** The entropy of shared belief of the search region changed after observation,  $FSR$  is short for *FlankSupportRange*,  $ROR$  is short for *RobotObservationRange*

as the human and intends to plan a time length identical with the predicted human motion length. We set the initial position by  $x_0 = y_0$  and the planning time length to  $D$ . These constraints yield a natural tree structure for the problem, which forms a tree by finding all “visitable” cells by (7). The problem is thus to find the sequence  $\mathbf{X} = [x^1; x^2; \dots x^D]$  of robot positions such that

$$\sum_{t=1}^D F(x^t) = H(bel^D) - H(bel^0) \tag{9}$$

is as large as possible. Performance can either be described as a summation of information gain at each time step or the total information gain reward. Putting this together with the constraint (7) from the team interaction model yields the constrained optimization problem for the team task model.

$$\begin{aligned} & \max_{x^1 \dots x^D} \sum_{t=1}^D F(x^t) \\ & \text{subject to } x^t \in N_{flank}^{human}(y^t) \cap N_{motion}^{robot}(x^{t-1}). \end{aligned} \tag{10}$$

Fig.4 shows a case of how the entropy of the shared belief of the search region has been updated by the search of the human-robot team. To visualize the entropy, we color the maximum value in black and minimum value in white, which determines the gray transition for the values in between. Before the search begins, we assume we have no

idea on the location of the search object so that all the cells have been colored black which shows the largest uncertainty. For visited cells, the entropy is reduced to zero. This means the entropy of the shared belief of this cell has been reduced to zero so that this cell has been colored white.

## 5.4 Simulation

Fig.4(a) shows how the entropy of the shared belief of the search region has been changed by a human search agent. The robot's teammate model assumes, via Equation (6), that the human has imperfect observation in neighboring cells, so the entropy of the believes on neighboring cells has been reduced less than the visited cells. The value of the gray color is determined by how the observation model is defined.

Based on the human path and how the shared belief of the environment has changed, the robot wingman plans a path to optimize the team information gain, subject to the teammate interaction model via Equation (7). Fig.4(b), 4(c) and 4(d) show the entropy of the shared belief of the search region after the search of the robot wingman within different parameters. We use arrows to label the planned path of the robot wingman. We can see that increasing the observation range of the robot will usually not influence the planned path for the robot wingman, as Fig.4(b) and 4(c) have the same path shapes. However, increasing the flank support range, which gives more motion freedom to the robot wingman, will lead to a new generated path, as shown in Fig.4(d).

## 6 Conclusion and Future Work

Based on shared mental model and search theory, we model team-based search as an information-based task using a Bayesian approach. A wingman robot has been introduced for this problem, with robot decision algorithms designed to support collaborative human-robot interaction. Using the entropy of the belief of the search region as a way of information measurement, we illustrate how the robot wingman will do path planning for collaborating with the human as an optimization problem. Using a specific case, we illustrate how the human robot collaboration on a search task will change the entropy of the belief of the search region, which works as a shared model on the environment from the team perspective. Here we only use a depth-first exhaustive search to find the optimal solution for wingman path planning. Future work will be focused on proposing an efficient and applicable solution for wingman path planning. Moreover, we will add more features on modeling the search environment, like obstacles and stochastic dynamics. Finally, we will relate problem modeling assumptions to the requirement of a shared mental model.

**Acknowledgments.** We would like to thank the U.S. Army Research Lab for providing funding for this work. The views of this paper do not necessarily reflect the views of the funding agency.

## References

- [1] Parasuraman, R., Sheridan, T., Wickens, C.: A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 30(3), 286–297 (2000)
- [2] Breazeal, C., Gray, J., Hoffman, G., Berlin, M.: Social robots: beyond tools to partners. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2004, pp. 551–556 (September 2004)
- [3] Hoffman, G., Breazeal, C.: Collaboration in human-robot teams. In: Proc. of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA (2004)
- [4] Salas, E., Fiore, S., Letsky, M.: Theories of team cognition: Cross-disciplinary perspectives. Routledge Academic (2011)
- [5] Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: The influence of shared mental models on team process and performance. *Journal of Applied Psychology* 85(2), 273 (2000)
- [6] Fiore, S.: Personal communication
- [7] Neerinx, M., de Greef, T., Smets, N., Sam, M.: Shared mental models of distributed human-robot teams for coordinated disaster responses. In: 2011 AAAI Fall Symposium Series (2011)
- [8] Roscheck, M., Goodrich, M.: Detection likelihood maps for wilderness search and rescue. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 327–332. IEEE (2012)
- [9] Casper, J., Murphy, R.: Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 33(3), 367–385 (2003)
- [10] Doroodgar, B., Ficocelli, M., Mobedi, B., Nejat, G.: The search for survivors: Cooperative human-robot interaction in search and rescue environments using semi-autonomous robots. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 2858–2863 (May 2010)
- [11] Ruangpayoongsak, N., Roth, H., Chudoba, J.: Mobile robots for search and rescue. In: 2005 IEEE International Safety, Security and Rescue Robotics, Workshop, pp. 212–217. IEEE (2005)
- [12] Burion, S.: Human detection for robotic urban search and rescue. Diploma Work (2004)
- [13] Nourbakhsh, I.R., Sycara, K., Koes, M., Yong, M., Lewis, M., Burion, S.: Human-robot teaming for search and rescue. *IEEE Pervasive Computing* 4(1), 72–79 (2005)
- [14] Burke, J., Murphy, R.: Human-robot interaction in usar technical search: two heads are better than one. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2004, pp. 307–312 (September 2004)
- [15] Thrun, S., Burgard, W., Fox, D., et al.: Probabilistic robotics, vol. 1. MIT Press, Cambridge (2005)
- [16] Cover, T., Thomas, J.: Elements of information theory. Wiley Interscience (2006)

# Assessing Interfaces Supporting Situational Awareness in Multi-agent Command and Control Tasks

Donald Kalar<sup>1,2</sup> and Collin Green<sup>2</sup>

<sup>1</sup> Department of Psychology

San Jose State University, San Jose CA, USA

<sup>2</sup> Human Systems Integration Division

NASA Ames Research Center, Moffett Field, CA, USA

donald.j.kalar@nasa.gov, collin.b.green@nasa.gov

**Abstract.** Here, we describe our efforts to uncover design principles for multi-agent supervision, command, and control by using real-time strategy (RTS) video games as a source of data and an experimental platform. Previously, we have argued that RTS games are an appropriate analog for multi-agent command and control [3] and that publicly-available data from gaming tournaments can be mined and analyzed to investigate human performance in such tasks [5]. We outline additional results produced by mining public game data and describe our first foray into using RTS games as an experimental platform where game actions (e.g., clicks, commands) are logged and integrated with eye-tracking data (e.g., saccades, fixations) to provide a more complete picture of human performance and a means to assess user interfaces for multi-agent command and control. We discuss the potential for this method to inform UI design and analysis for these and other tasks.

**Keywords:** Situation Awareness, Automation, RTS, Gaze Tracking, User Interfaces.

## 1 Investigating Multi-agent Command and Control Using Real-Time Strategy Games

Robotic sensor and on-board computing capabilities are advancing at an accelerated pace, enabling an increasing number of scenarios where robots can be deployed as autonomous or semi-autonomous agents rather than needing to be controlled via closed-loop tele-operation. These transitions allow the human operator to take on the role of commander or supervisor over multiple agents and increases the overall task execution rate for a fixed number of human users. Effective supervisory control systems require user interfaces that are compatible with the task demands and limitations of human memory, perception, cognition, attention, and actions. Optimized systems will maximize task execution while minimizing errors, allowing the human to command and control a greater number of autonomous agents at an acceptable level of performance.

One approach to developing candidate interfaces to support these multi-agent command and control scenarios is to investigate other domains that have similar or overlapping task characteristics [3]. This allows for more informed decisions when building interfaces, and can help to identify complications or limitations before building and testing high-fidelity systems. Real-Time Strategy (RTS) video games are one analog domain that can inform many aspects of how a human operator commands and controls multiple semi-autonomous agents in a dynamic environment under constraints. RTS games, as they are commonly played, constitute relatively a complex multi-agent supervision and control problem: A player will command ten to 200 units at a time, including as many as 20 different types with different characteristics and unique capabilities. The units will be deployed in service of several overarching strategic tasks (e.g., resource collection, construction, defense, scouting, offense) with multiple complex subtasks composing or supporting each task. To enable players to successfully manage the game (and even enjoy doing so) game user interface (UI) designers have—through careful design, evaluation, user research, and iteration—built RTS game UIs that are effective in this context.

Can the designs and design principles that are effective in RTS games be applied to supervisory control of multiple robots? Some evidence suggests that they can: at least one robotic control interface has been built to mimic a conventional RTS game and shown to be effective for robotic command and control [4]. However, merely duplicating interface and interaction designs from RTS games is not an adequate basis for building interfaces for all multi-agent command and control tasks. RTS games are distinctly *dissimilar* to real robotic control tasks in that RTS games do not include complexities like uncertainty and latency, both of which can have important consequences for human-robotic interaction (e.g., see [2]). Yet, RTS games can provide some design inspiration for robotic command and control interfaces and they also constitute a good platform for investigating the how interfaces, interactions, and properties of human performance and cognition may shape multi-robot control.

Like many real-world multi-agent tasks, these games demand that their players maintain high levels of situational awareness (SA) throughout the entire course of the game. Players must maintain a basic economy by acquiring and managing raw materials, monitor unit construction, explore terrain to monitor their opponent and environment, all while staging offensive attacks and defending their own bases and resources. Despite the high cognitive demand of these games, devoted players develop high levels of expertise, not unlike professional athletes. Video games and video game players are also of interest due to findings suggesting that game play may boost cognitive function, though these findings are still an open matter for investigation [1].

In summary, there is reason to believe that RTS games may be useful to researchers in human-robotic interaction inasmuch as they provide a platform for evaluating interface design and for assess human performance in a context that is similar (though not identical) to supervisory control of multiple robotic agents. In the remainder of this paper we outline our progress in this direction.

We have used publicly-available game archives to obtain data that help to shape and refine hypotheses about supervisory control and situational awareness, and we have recently begun collecting enhanced game data (user game actions plus user eye movement data) in an effort to better understand multi-robot HRI.

## 2 Mining and Analysis of Public RTS Game Archives

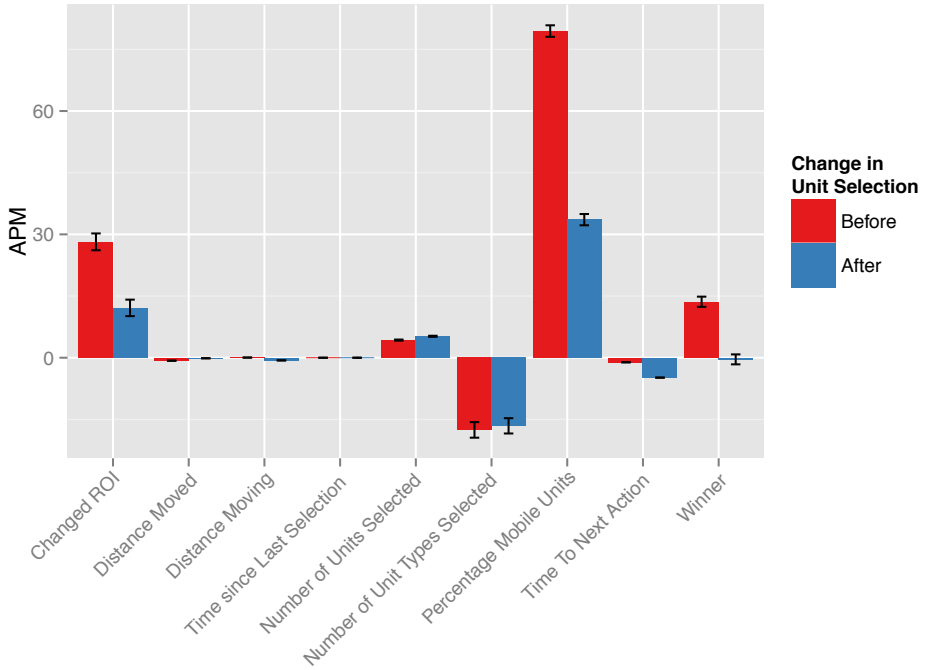
RTS Games such as *Starcraft II* often generate ‘replay’ files containing a timestamped log of every player input for later playback. These replay files are routinely posted online for general viewing, and it is especially common for expert-level tournaments to make these match replay files available. This provides a large and rich dataset for meta-analysis. In previous work, we analyzed 220 *Starcraft II* matches from the Championship Bracket at the Major League Gaming Pro Circuit 2011 in Orlando [5]. We investigated how different factors impacted instantaneous changes in player Actions Per Minute (APM) whenever a new unit or set of units were selected by the player in the game (see Figure 1). APM was chosen because is the standard heuristic for quantifying the skill-level of a player, and has been consistently correlated with winning games [6]. Expert players are able to maintain a rate of hundreds of APM throughout an entire match.

While the data are very noisy, the huge datasets allowed us to extract several small but interesting signals. For example, euclidean distance between sequential unit selections was not predictive of changes in APM, but changes between spatial clusters (ROIs) of actions did show an increase in APM. Players were faster to act when selecting larger groups of units, but slower to act if those groups were heterogeneous (i.e., multiple unit types in a single selected group). Players were much faster in issuing commands to mobile units when compared to stationary units. As expected, the general trend was that winners had greater APM than losers, but interestingly that increase in APM was only significantly different from zero immediately prior to a unit selection, not after selecting a new unit.

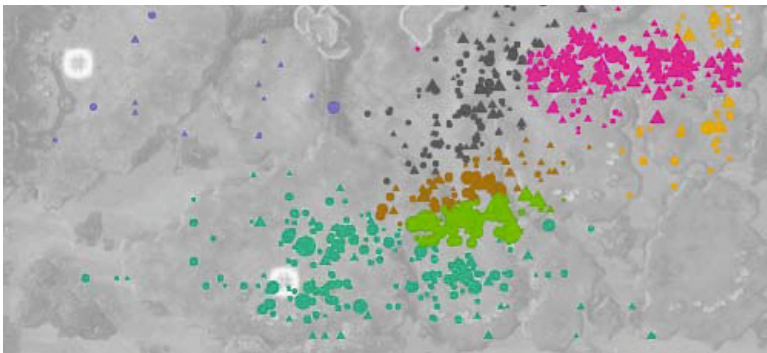
One major limitation to this approach is that, while these findings provide small hints as to how the players are interacting with the game interface, these data are necessarily ambiguous. A player can select a unit in order to gain SA (selecting allows the user to inspect the unit’s current status), or the selection may be part of a series of commands based on existing SA.

## 3 RTS Games as a Platform for HRI Experiments

In order to better understand how a player uses the interface to support game play, we are recruiting participants of varying skill level to play the game while having their game actions logged and their eye movements tracked and recorded. The user interface provides a number of controls to support various aspects of SA, as well as interfaces to support command execution. These controls are spatially distinct, so we can know how frequent and how long players of varying



**Fig. 1.** Regression model output predicting changes in instantaneous Actions Per Minute (APM) as a result of different game transitions or states. Red bars indicate the estimated APM using a one-second window prior to unit selection; blue bars are the estimated APM using a one-second after selection window.



**Fig. 2.** An example of regions of interest (ROIs) generated from player actions during a game. Each color denotes a different ROI, as extracted by a standard expectation-maximization clustering algorithm. There were observed changes in player action rate when transitioning between ROIs, but no observed differences for action rate based solely on Euclidean distance (see Figure 1).

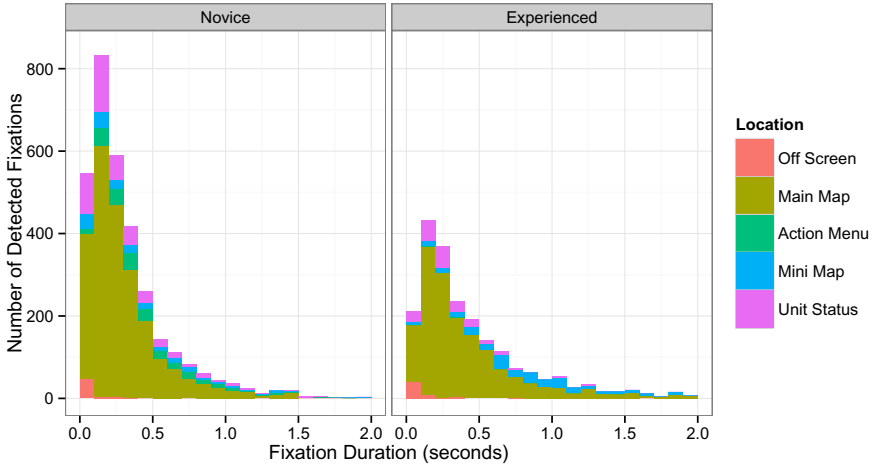
skill level scan away from the main view (the game map) to monitor unit status, game resources, regions that are outside of view from the main display, and command interfaces (see Figure 3). We predict qualitative differences between more expert and novice players in how frequently they sample these supporting controls and the total length of time they fixate on these controls rather than the main display.



**Fig. 3.** An annotated UI screenshot. The user is presented with a main overhead display of units in the environment (top). A mini map (bottom, left) provides a condensed view of the entire gameplay area, which is too large to view in the main display. A unit status window (bottom, center) provides information about currently selected units in the game, including details such as consumables status and task progress. The action menu (bottom right) provides a set of buttons to issue commands to the selected units, dynamically changing the set of buttons available based on what actions are available to issue the current selection.

It is worth noting that the addition of eye movement data collection to game logging may be especially useful for understanding how situational awareness (SA) is acquired and maintained. RTS game UIs (like many other UIs) conflate the acquisition of SA and the use of SA in issuing commands: the selection of a unit in an RTS game might be intended to reveal that unit's identity and status (acquire/update SA), to enable the issuance of a new command to that unit (apply SA), or both. Game log data only reveal that a unit was selected and what command (if any) was issued to that unit. However, eye movement data can reveal whether unit selection is followed by sampling of unit status information, examination of command options, or other eye movements associated with SA acquisition or SA use.

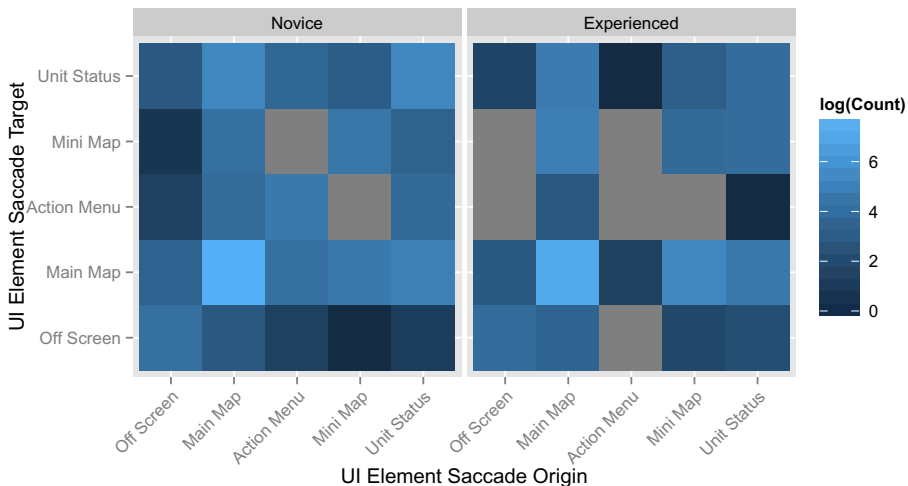




**Fig. 4.** Measured fixations for a single game. Gaze fixation locations and durations are plotted in this stacked histogram for both a novice and an experienced player for a single game. While the total observed fixation time during the games are almost identical (both just over seventeen minutes of detected gaze), the experienced player tends to fixate for longer durations while the novice player scans the UI at a much faster rate. The experienced player also effectively ignores elements of the UI (i.e., the Action Menu), while the novice is still reliant on the information in that display.

## 4 Future Experiments with RTS Games

Our initial explorations in RTS game data are promising. A number of compelling hypotheses have been generated from the analysis of one sample of RTS game archive data and our pilot experiment with just two users is already showing great promise for quantifying differences in the play of novice and expert users (c.f., left and right plots, Figures 4 & 5). The combination of a rich and engaging user task with high-fidelity human performance data gather makes the RTS game experiment platform a good choice for further investigations of HRI, generally. For example, some of the dimensions upon which a multi-agent control task might vary are easily explored in an RTS game experiment. As a candidate study, in the near future we hope to use this platform to examine how the number and diversity of units affects human performance in simple tasks like guiding a team of units between points on a map. RTS games (and game-editing utilities built into them) make it possible to vary the terrain, the size of the robot team, and the relative homogeneity or heterogeneity of the team. Metrics such as task time, path efficiency, or avoidance of hazards (i.e. enemies in an RTS game) can be used to evaluate task performance, and game log and eye movement data can reveal differences in supervision and control strategies for different users and different robot team compositions.



**Fig. 5.** Saccade origin and target frequency by UI element for a novice and an experienced player, log transformed. While the vast majority of saccades both originate and terminate in the Main Map region of the UI, there are other interesting scan patterns in the data. The novice spends more time scanning off of the display screen than the experienced player, presumably due to a difference in familiarity with keyboard issued commands. The experienced player also very rarely ever scans to the Action menu (as also illustrated in Figure 4), as that information can be determined from game state. Neither the novice nor the expert scan between the Mini Map and Action Menu UI elements. This makes sense, as the information and interactions provided in those controls do not directly relate to each other, regardless of the sophistication of the player.

Another issue that may be amenable to investigation with this approach concerns multi-agent command and control as one component of a complex environment involving several tasks. The UI and user strategies that best support multi-agent supervision when performed as an isolated task may not be the same as those that best support it when performed concurrently with another task (like communicating with another human). Asking participants to play RTS games in multi-tasking experiments may aid us in understanding how control interfaces should be altered for busy workplaces.

A related issue concerns collaborative control and human-human communication. RTS game support multiplayer matches where a two or more players can control teams of units in a single environment and can act in a collaborative or adversarial (or ambiguous) manner with other human or computer players. Future experiments may investigate verbal (or other) communication patterns among human players and may also reveal changes in how SA is acquired and maintained when multiple humans operate robotic teams in a shared environment.

In general, it seems promising to exploit the similarity between RTS games and multi-agent command and control even though these are not strictly identical

domains. The availability of archive data, the existence of an population of expert users, and the ease of implementing interesting experiments all support fast and effective data collection. The resulting data will inform our understanding of interface design, our understanding of human performance and strategy, and our understudying of situational awareness in complex tasks.

## References

1. Walter, R., Boot, D.P.: Blackely, and Daniel J Simons. Do action video games improve perception and cognition? *Frontiers in Psychology* 2 (2011)
2. Ellis, S.R., Yeom, K., Adelstein, B.D.: Human control in rotated frames: anisotropies in the misalignment disturbance function of pitch, roll, and yaw. In: *Proceedings of the Human Factors and Ergonomics Society* (2012)
3. Green, C., Kalar, D.: Design inspiration for interactions between humans and multiple robots. Presentation at the 4th IEEE Conference on Space Mission Challenges for IT (2011)
4. Jones, H.L., Snyder, M.: Supervisory control of multiple robots based on a real-time strategy game interaction paradigm. In: *Proc. of the 2001 IEEE International Conference on Systems, Man, and Cybernetics* (2001)
5. Kalar, D., Green, C.: Understanding situational awareness in multi-unit supervisory control through data-mining and modeling with real-time strategy games. In: *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE (2012)
6. Lewis, J.M., Trinh, P., Kirsh, D.: A corpus analysis of strategy video game play in *Starcraft: Brood War*. In: Miyake, N., Peebles, D., Cooper, R.P. (eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pp. 671–676 (2011)

# Cognitive Models of Decision Making Processes for Human-Robot Interaction

Christian Lebiere<sup>1</sup>, Florian Jentsch<sup>2</sup>, and Scott Ososky<sup>2</sup>

<sup>1</sup> Psychology Department, Carnegie Mellon University  
5000 Forbes Avenue, Pittsburgh PA 15208  
cl@cmu.edu

<sup>2</sup> Institute for Simulation & Training, University of Central Florida  
4000 Central Florida Blvd, Orlando, FL 32816  
Florian.Jentsch@ucf.edu, sososky@ist.ucf.edu

**Abstract.** A fundamental aspect of human-robot interaction is the ability to generate expectations for the decisions of one's teammate(s) in order to coordinate plans of actions. Cognitive models provide a promising approach by allowing both a robot to model a human teammate's decision process as well as by modeling the process through which a human develops expectations regarding its robot partner's actions. We describe a general cognitive model developed using the ACT-R cognitive architecture that can apply to any situation that could be formalized using decision trees expressed in the form of instructions for the model to execute. The model is composed of three general components: instructions on how to perform the task, situational knowledge, and past decision instances. The model is trained using decision instances from a human expert, and its performance is compared to that of the expert.

**Keywords:** Human-robot interaction, shared mental models, cognitive modeling, cognitive architectures, ACT-R, decision trees.

## 1 Introduction

A fundamental aspect of human-robot interaction is the ability to generate expectations for the decisions of one's teammate(s) in order to coordinate plans of actions. Shared mental models of the knowledge and decision procedures of human and robotic teammates enable them to act as a team rather than a collection of individuals that have to be explicitly controlled [1]. Mental models include a representation of the current situation, the various entities (humans, robots, even animals) involved, their capacities and limitations, and some decisions of their past decisions and actions. Current mental models are often non-computational descriptions of that information that only provide a qualitative understanding and limited predictive power.

Computational cognitive models provide a promising approach to that problem. Cognitive models can provide a computational link from the shared mental model literature to the domain of robotic control and intelligence. By computationally representing the cognitive processes and representations underlying shared mental

models, they enable them to be leveraged in a number of ways to improve human-robot interaction. First, they provide a quantitative, predictive understanding of human shared mental models and their impact on team performance. Conversely, they also provide a cognitively based computational basis for the implementation of mental models in robots that are similar to those of human teammates. Thus, they foster better teamwork by allowing human and robotic teammates to work in similar ways by representing and simulating the internal processes of the others. Finally, they support the improved design of human-robot interaction tools and protocols by enabling quantitative predictions of their effectiveness.

Mental models can be computationally represented in a number of ways. They can consist of an ontology of concepts and decisions that represent the factors involved in shared mental models. They can take the form of symbolic frameworks such as decision trees and semantic networks that capture the decision procedures followed by teammates. They can also be represented using statistical frameworks such as Bayesian networks or Markov models that capture the statistical regularities of the environment and decisions made.

Cognitive models provide a direct computational instantiation of the representations and mechanisms involved in shared mental models. They provide an account of situation awareness by mapping it to the representation of the current situation and its evolution in the various memory stores (working, short-term, episodic, long-term semantic) available. They include an account of perceptual-motor limitations such as attentional processes. Finally, the basic cognitive processes themselves combine capabilities such as associative pattern-matching and adaptivity with capacity limitations such as limited working memory, memory decay, and stochasticity.

Cognitive models have been used for a number of purposes related to shared mental models. A methodology called instance-based learning has been developed to learn to control complex systems by observing the actions of another controller and making decisions that generalize from those observations [2]. Perspective-taking in spatial domains such as hide-and-seek or collaborative work in space has been used to infer the knowledge of the current situation on the part of another human or robotic entity [3]. Predicting decisions of other entities has been implemented using a number of cognitive processes including theory of mind recursion [4], imagery-based simulation [5] and memory retrieval [6]. Cognitive models have been used to implement the execution of joint plans of actions between synthetic teammates in virtual and robotic simulations [7].

In this paper, we introduce a joint pursuit task involving shared mental models between human and robotic teammates. We describe a general cognitive model developed using the ACT-R cognitive architecture that can apply to any situation that could be formalized using decision trees expressed in the form of instructions for the model to execute. The model is composed of three general components: instructions on how to perform the task, situational knowledge, and past decision instances. The model is trained using decision instances from a human expert, and its performance is compared to that of the expert. Finally we discuss open issues and further work planned in the development of our cognitive representation of shared mental models.

## 2 Pursuit Scenario

A foot pursuit scenario was developed that highlights the need for close cooperation and an understanding of the capabilities and vulnerabilities of the various human and robotic teammates involved. A decision structure for the pursuit scenario was drafted based on publicly available information on police foot pursuit procedures, online video, and incident reports. These were then altered to the appropriateness of a robot attempting to make a similar action selection about whether or not to pursue a suspect.

ID	A	B	C	D	E	F	G
Model Structure					Scenario 1	Scenario 2	Scenario 3
ID	Item	Model Information	Data Type	Unit			
1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2	2.0	2.0	2.0	2.0	2.0	2.0	2.0
3	3.0	3.0	3.0	3.0	3.0	3.0	3.0
4	4.0	4.0	4.0	4.0	4.0	4.0	4.0
5	5.0	5.0	5.0	5.0	5.0	5.0	5.0
6	6.0	6.0	6.0	6.0	6.0	6.0	6.0
7	7.0	7.0	7.0	7.0	7.0	7.0	7.0
8	8.0	8.0	8.0	8.0	8.0	8.0	8.0
9	9.0	9.0	9.0	9.0	9.0	9.0	9.0
10	10.0	10.0	10.0	10.0	10.0	10.0	10.0
11	11.0	11.0	11.0	11.0	11.0	11.0	11.0
12	12.0	12.0	12.0	12.0	12.0	12.0	12.0
13	13.0	13.0	13.0	13.0	13.0	13.0	13.0
14	14.0	14.0	14.0	14.0	14.0	14.0	14.0
15	15.0	15.0	15.0	15.0	15.0	15.0	15.0
16	16.0	16.0	16.0	16.0	16.0	16.0	16.0
17	17.0	17.0	17.0	17.0	17.0	17.0	17.0
18	18.0	18.0	18.0	18.0	18.0	18.0	18.0
19	19.0	19.0	19.0	19.0	19.0	19.0	19.0
20	20.0	20.0	20.0	20.0	20.0	20.0	20.0
21	21.0	21.0	21.0	21.0	21.0	21.0	21.0
22	22.0	22.0	22.0	22.0	22.0	22.0	22.0
23	23.0	23.0	23.0	23.0	23.0	23.0	23.0
24	24.0	24.0	24.0	24.0	24.0	24.0	24.0
25	25.0	25.0	25.0	25.0	25.0	25.0	25.0
26	26.0	26.0	26.0	26.0	26.0	26.0	26.0
27	27.0	27.0	27.0	27.0	27.0	27.0	27.0
28	28.0	28.0	28.0	28.0	28.0	28.0	28.0
29	29.0	29.0	29.0	29.0	29.0	29.0	29.0
30	30.0	30.0	30.0	30.0	30.0	30.0	30.0
31	31.0	31.0	31.0	31.0	31.0	31.0	31.0
32	32.0	32.0	32.0	32.0	32.0	32.0	32.0
33	33.0	33.0	33.0	33.0	33.0	33.0	33.0
34	34.0	34.0	34.0	34.0	34.0	34.0	34.0
35	35.0	35.0	35.0	35.0	35.0	35.0	35.0
36	36.0	36.0	36.0	36.0	36.0	36.0	36.0
37	37.0	37.0	37.0	37.0	37.0	37.0	37.0
38	38.0	38.0	38.0	38.0	38.0	38.0	38.0
39	39.0	39.0	39.0	39.0	39.0	39.0	39.0
40	40.0	40.0	40.0	40.0	40.0	40.0	40.0
41	41.0	41.0	41.0	41.0	41.0	41.0	41.0
42	42.0	42.0	42.0	42.0	42.0	42.0	42.0
43	43.0	43.0	43.0	43.0	43.0	43.0	43.0
44	44.0	44.0	44.0	44.0	44.0	44.0	44.0
45	45.0	45.0	45.0	45.0	45.0	45.0	45.0
46	46.0	46.0	46.0	46.0	46.0	46.0	46.0
47	47.0	47.0	47.0	47.0	47.0	47.0	47.0
48	48.0	48.0	48.0	48.0	48.0	48.0	48.0
49	49.0	49.0	49.0	49.0	49.0	49.0	49.0
50	50.0	50.0	50.0	50.0	50.0	50.0	50.0
51	51.0	51.0	51.0	51.0	51.0	51.0	51.0
52	52.0	52.0	52.0	52.0	52.0	52.0	52.0
53	53.0	53.0	53.0	53.0	53.0	53.0	53.0
54	54.0	54.0	54.0	54.0	54.0	54.0	54.0
55	55.0	55.0	55.0	55.0	55.0	55.0	55.0
56	56.0	56.0	56.0	56.0	56.0	56.0	56.0
57	57.0	57.0	57.0	57.0	57.0	57.0	57.0
58	58.0	58.0	58.0	58.0	58.0	58.0	58.0
59	59.0	59.0	59.0	59.0	59.0	59.0	59.0
60	60.0	60.0	60.0	60.0	60.0	60.0	60.0
61	61.0	61.0	61.0	61.0	61.0	61.0	61.0
62	62.0	62.0	62.0	62.0	62.0	62.0	62.0
63	63.0	63.0	63.0	63.0	63.0	63.0	63.0
64	64.0	64.0	64.0	64.0	64.0	64.0	64.0
65	65.0	65.0	65.0	65.0	65.0	65.0	65.0
66	66.0	66.0	66.0	66.0	66.0	66.0	66.0
67	67.0	67.0	67.0	67.0	67.0	67.0	67.0
68	68.0	68.0	68.0	68.0	68.0	68.0	68.0
69	69.0	69.0	69.0	69.0	69.0	69.0	69.0
70	70.0	70.0	70.0	70.0	70.0	70.0	70.0
71	71.0	71.0	71.0	71.0	71.0	71.0	71.0
72	72.0	72.0	72.0	72.0	72.0	72.0	72.0
73	73.0	73.0	73.0	73.0	73.0	73.0	73.0
74	74.0	74.0	74.0	74.0	74.0	74.0	74.0
75	75.0	75.0	75.0	75.0	75.0	75.0	75.0
76	76.0	76.0	76.0	76.0	76.0	76.0	76.0
77	77.0	77.0	77.0	77.0	77.0	77.0	77.0
78	78.0	78.0	78.0	78.0	78.0	78.0	78.0
79	79.0	79.0	79.0	79.0	79.0	79.0	79.0
80	80.0	80.0	80.0	80.0	80.0	80.0	80.0
81	81.0	81.0	81.0	81.0	81.0	81.0	81.0
82	82.0	82.0	82.0	82.0	82.0	82.0	82.0
83	83.0	83.0	83.0	83.0	83.0	83.0	83.0
84	84.0	84.0	84.0	84.0	84.0	84.0	84.0
85	85.0	85.0	85.0	85.0	85.0	85.0	85.0
86	86.0	86.0	86.0	86.0	86.0	86.0	86.0
87	87.0	87.0	87.0	87.0	87.0	87.0	87.0
88	88.0	88.0	88.0	88.0	88.0	88.0	88.0
89	89.0	89.0	89.0	89.0	89.0	89.0	89.0
90	90.0	90.0	90.0	90.0	90.0	90.0	90.0
91	91.0	91.0	91.0	91.0	91.0	91.0	91.0
92	92.0	92.0	92.0	92.0	92.0	92.0	92.0
93	93.0	93.0	93.0	93.0	93.0	93.0	93.0
94	94.0	94.0	94.0	94.0	94.0	94.0	94.0
95	95.0	95.0	95.0	95.0	95.0	95.0	95.0
96	96.0	96.0	96.0	96.0	96.0	96.0	96.0
97	97.0	97.0	97.0	97.0	97.0	97.0	97.0
98	98.0	98.0	98.0	98.0	98.0	98.0	98.0
99	99.0	99.0	99.0	99.0	99.0	99.0	99.0
100	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Fig. 1. Representation of the foot pursuit scenario

Eight scenarios were developed representing different situations and resulting decisions. Data related to the scenarios were encoded and organized according to shared mental models held by expert teams, including item categories such as equipment, task, team interaction and team characteristics (Fig. 1). A decision tree was then built using information from police foot pursuit procedures leading to four possible initial decisions: soldier-only pursuit, robot-only pursuit, joint soldier-robot team pursuit, and holding position without pursuit and reporting the incident (Fig. 2).

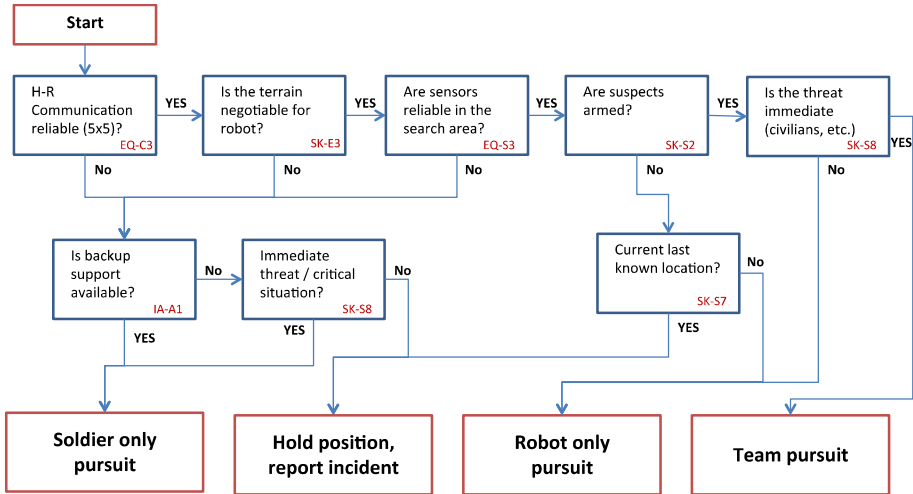


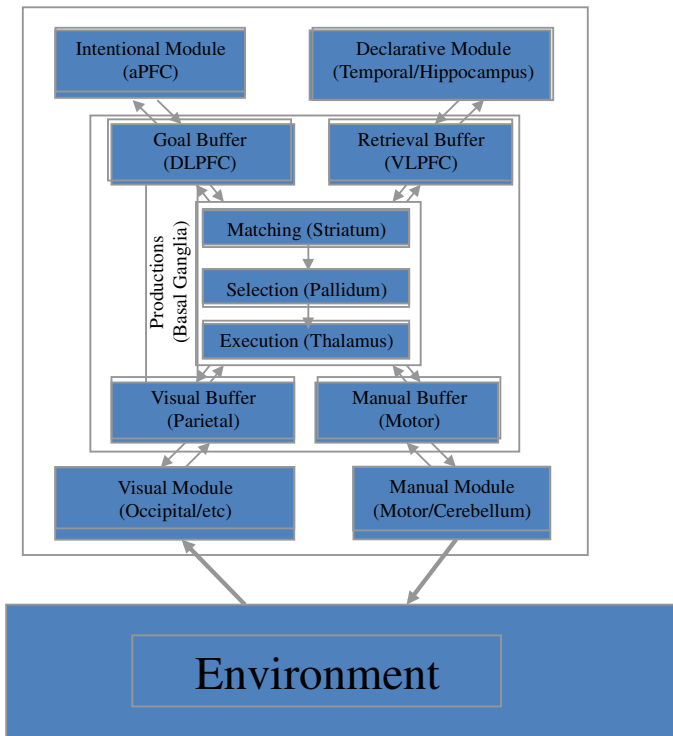
Fig. 2. Decision tree leading to four distinct actions

For each decision, the most critical item was listed, although other factors were usually also considered in the decision. The scenario data provided information about both the items involved in each decision and the correct decision according to a human expert. Finally, the human expert rank-ordered the four possible actions according to their suitability for each of the eight scenarios.

### 3 Cognitive Model

The cognitive model was developed using the ACT-R cognitive architecture [8], [9]. Cognitive architectures are computational representation of invariant cognitive mechanisms specified by unified theories of cognition. ACT-R is a modular architecture, reflecting neural constraints, composed of asynchronous modules coordinated through a central procedural system (Fig. 3). The procedural system is in charge of behavior selection and more generally the synchronization of the flow of information between the other modules. It is implemented as a production system where competing production rules are selected based on their utilities, learning through a reinforcement mechanism from the rewards and costs associated with their actions. The production system conditions are matched against limited-capacity buffers that control the interaction with the other modules by enabling a single command (e.g.,

retrieval of information, focus of visual attention) to be given at a time to a given module, and a single result to be returned (e.g., chunk retrieved from memory, visual item encoded). A declarative memory module holds both short-term information, such as the details of the current situation, as well as long-term knowledge, such as the procedural rules to follow. Access to memory is controlled by an activation calculus that determines the availability of chunks of information according to their history of use such as recency, frequency, and degree of semantic match. Learning mechanisms control both the automatic acquisition of symbolic structures such as production rules and declarative chunks, and the tuning of their subsymbolic parameters (utility and activation) to the structure of the environment. The perceptual-motor modules reflect human factor limitations such as attentional bottlenecks. Individual differences can be represented both in terms of differences in procedural skills and declarative knowledge, as well as in terms of architectural parameters controlling basic cognitive processes such as spreading of activation.



**Fig. 3.** ACT-R Cognitive Architecture

As is standard with cognitive modeling practice, we initially considered developing a cognitive model of this specific domain and decision procedure. Instead, we aimed to develop a general cognitive model of shared mental models independent of any domain or decision procedure. Therefore, we generalized our initial cognitive model



to apply beyond the current scenario to any situation that could be formalized using a decision tree procedure. This general ACT-R model takes mental models formalized as decision trees and expresses them in the form of instructions for the model to execute. The declarative knowledge of the model is composed of three general components.

The first part is the general declarative knowledge, encoded as instructions, of how to carry out decisions in a type of situation. Each decision is represented as a set of chunks that represent the situation variables relevant to the decision as well as how to chain multiple decisions together in an overall decision tree. General knowledge on how to carry out the decision, such as heuristic rules, could also be represented but is not currently. This first part can be seen as the core mental model relevant to that situation. At this point we do not differentiate between mental models for different individuals, including teammates and/or opponents, assuming that those models are generally shared between individuals. However, elaborating the representation to allow for mental models of individual decision-making would be quite possible.

The second part of the model is the declarative representation of specific situational knowledge, both the current one as well as past ones. This knowledge is composed of one chunk for each situational variable and its value, for each given situation. It also encodes instances of decisions made in the context of past situations, which are generalized to make decisions in future situations. This part of the model can be described as the situation awareness of the situation. Again, as for the mental model above, this representation of the situation is assumed shared between individuals but could be individualized as well.

The third part of the model is the representation of past instances of decisions. In keeping with the instance-based learning methodology, those decisions consist of a set of the type of decision in the decision tree, a set of values describing the relevant item in a past scenario, and the decision that was taken. Those instances are learned from the human expert annotation of the scenarios provided and represented his individual expertise. It is quite straightforward for models to acquire experience from different experts (or a combination of them) thus allowing for an individualized style of decisions tailored to specific teammates. There is no hardcoded decision logic: each decision depends on matching against past instances involving the subsymbolic (statistical) level of the architecture, with activation processes reflecting factors such as recency, frequency and degree of partial matching as well as stochastic factors resulting in probabilistic decisions.

The final part of the model is the procedural logic that controls how the first part of the model, the shared mental model of procedures, is applied to the second, the situation awareness of the situation, using the third, the experience with similar situations, to generate a series of decisions in the current situation. While the other parts of the model can be quite complex and specific and require large numbers of chunks to be represented, this part of the model is quite general, being applicable to any mental model and situation that can be expressed in the current decision tree format, and compact, being represented using only 7 production rules. The production rules learn to retrieve and execute the instructions to interpret the mental model. Each decision is represented as sequence of chained steps to retrieve and encode in working memory

the various pieces of information from the current situation relevant to the decision. The model then makes a decision by retrieving and generalizing past decision instances for similar situations. Each decision leads to another according to the logic encoded in the decision tree procedures, until a decision regarding one of the four possible actions is taken.

## 4 Results

Fig. 4 illustrates how the model can learn to make decisions without being given explicit logic but instead individual decision instances. Instructions represented the factors involved in the decision (abstracted here as Factor1 and Factor2). Reflecting the probabilistic nature of the decisions, the model was run in Monte Carlo mode to produce a distribution of decision probabilities. The model has inferred that Factor1 (a binary yes/no item) results in a fairly constant probability increase of a “yes” decision while Factor2 (a range item with integer values between zero and five) resulted in a gradual increase in “yes” decisions of about 50% across its range of values.

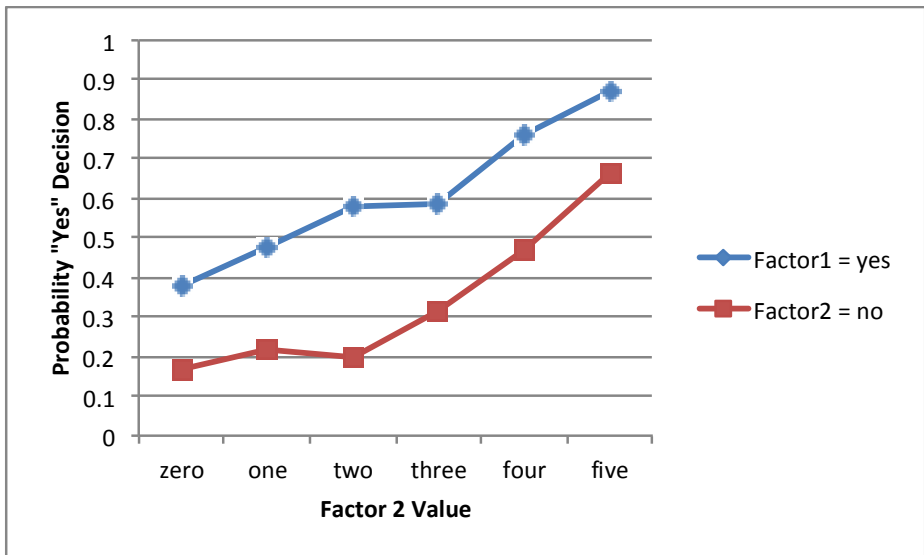
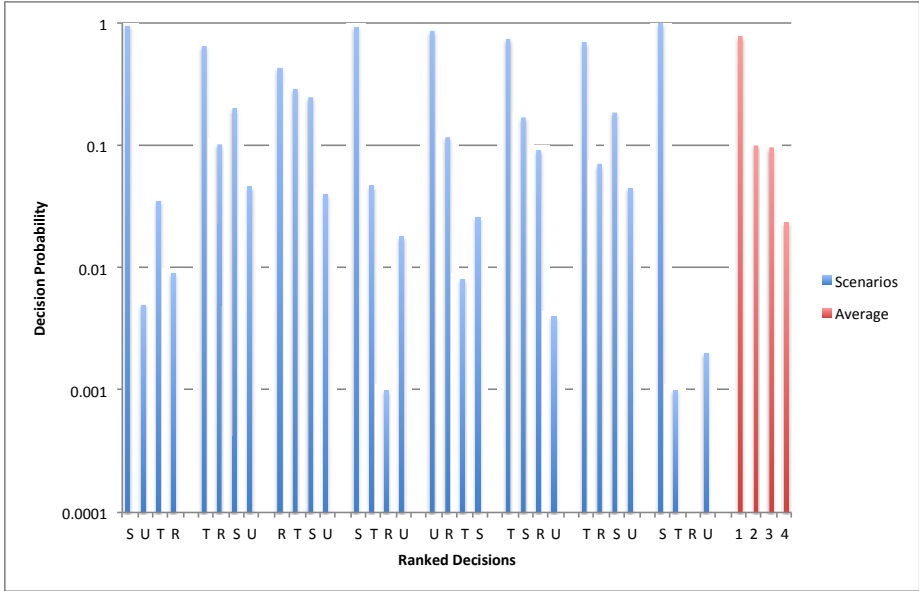


Fig. 4. Probability distribution for a two-factor decision

Since the decisions resulted from past instances rather than hardcoded rules, the model was trained on a subset of the scenarios. We cross-validated the model performance by selecting all subsets of 7 out of the 8 scenarios then tested its generalization to the remaining scenario. Model performance is summarized in Fig. 5, with decision probabilities plotted on a log scale. One can see that for each scenario, one of four possible actions usually (but not always, e.g. the third scenario) stands out as the correct one. Aggregating all decisions across scenarios according to the ranking (1 to 4) provided by the human expert, one can see that the cognitive model has

learned to reproduce the expert's judgment, overwhelmingly (~80%) favoring the #1-ranked decision. The decisions ranked #2 and #3 result in much closer probabilities of about 10% while the #4-ranked decision is almost never taken.



**Fig. 5.** Model performance compared to subject judgments. S means soldier pursuit, R means robot pursuit, T means robot and soldier pursue together, and U means abandon pursuit.

## 5 Discussion

While the existing model is quite general, a number of improvements are possible. Reflecting the general transition from declarative knowledge to procedural skills, the model could proceduralize the individual decision steps from declarative instructions to production rules to replicate the learning curve from novice performance to proficiency and expertise. ACT-R includes a production compilation mechanism that compiles retrievals from declarative memory into production rules to implement this learning process. A feature selection process using the utility learning mechanism could be added to encode and use only a subset of data items for each decision. Data items would only be encoded to the extent that they contributed to making the correct decision. Items that do not contribute to a correct decision would be dropped from the procedure, in a manner similar to that by which experts decide which aspects of a situation warrant their limited attention in high-pressure situations. Similarly, the model could learn shortcuts that combine multiple individual binary decisions into a single, multi-outcome decision, as when experts can recognize a situation and make a rapid, almost instantaneous decision instead of going through the same painstaking process that novices go through. This is possible in our model since the final

decisions involving the four possible actions can be represented in the same instance-based manner as the individual decisions leading to it. Finally, since Monte Carlo simulations are not a cognitively plausible process, it would be desirable to generate rankings of the various actions directly from the activations of the memory retrievals involved in the decision process.

Alternative implementations are also possible. Adopting a Bayesian network formalism (in keeping with the Bayesian underpinnings of the ACT-R activation calculus) would provide an alternative to decision trees in order to enhance generalization in multi-step decisions.

Improvements in the model validation are also possible. Model performance could be validated against human participants data along the entire learning curve, reflecting the learning processes currently existing in the model (strengthening of instructions, accumulation of decision instances) and those described above. As mentioned previously, the model could also be “seeded” with decision instances from different human experts, and its performance compared to that of the individual experts. Integrating the cognitive model in multi-agent simulations would permit to validate it in a dynamic decision-making setting in which a series of decisions is taken rather than a single one. Finally, and most fundamentally, integrating the cognitive model on a robotic platform would allow us to assess its ability to improve human-robot interaction through the computational implementation of shared mental models. This would involve adding to the model the procedural control to perform inferences about the situation awareness knowledge of other entities, and the procedural control to use mental models to plan joint actions involving both teammates and opponents.

**Acknowledgments.** This work was conducted through collaborative participation in the Robotics Consortium sponsored by the U.S Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement W911NF-10-2-0016.

## References

1. Jentsch, F., Ososky, S., Schuster, D., Fiore, S., Shumaker, R., Lebiere, C., Kurup, U., Oh, J., Stentz, A.: The Importance of Shared Mental Models and Shared Situation Awareness for Transforming Robots from Tools to Teammates. In: Proceedings of the 2012 SPIE Conference, Baltimore, MD (2012)
2. Gonzalez, C., Lerch, F.J., Lebiere, C.: Instance-based learning in dynamic decision making. *Cognitive Science* 27(4), 591–635 (2003)
3. Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E., Schultz, A.C.: Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 35(4), 460–470 (2005)
4. Hiatt, L.M., Trafton, J.G.: A cognitive model of theory of mind. In: Salvucci, D.D., Gunzelmann, G. (eds.) Proceedings of the 10th International Conference on Cognitive Modeling, pp. 91–96. Drexel University, Philadelphia (2010)
5. Wintermute, S.: Using Imagery to Simplify Perceptual Abstraction in Reinforcement Learning Agents. In: Proceedings of the 2010 AAAI Conference on Artificial Intelligence (2010)

6. West, R.L., Lebiere, C.: Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Journal of Cognitive Systems Research* 1(4), 221–239 (2001)
7. Best, B.J., Lebiere, C.: Cognitive agents interacting in real and virtual worlds. In: Sun, R. (ed.) *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge University Press, NY (2006)
8. Anderson, J.R., Lebiere, C.: *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Mahwah (1998)
9. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. *Psychological Review* 111(4), 1036–1060 (2004)

# Human Considerations in the Application of Cognitive Decision Models for HRI

Scott Ososky, Florian Jentsch, and Elizabeth Phillips

University of Central Florida, Partnership II, 3100 Technology Parkway, Orlando, FL 32826  
sososky@ist.ucf.edu, Florian.Jentsch@ucf.edu,  
ephillips@knights.ucf.edu

**Abstract.** In order for autonomous robots to succeed as useful teammates for humans, it is necessary to examine the lens through which human users view, understand, and predict robotic behavior and abilities. To further study this, we conducted an experiment in which participants viewed video segments of a robot in a task-oriented environment, and were asked to explain what the robot was doing, and would likely do next. Results showed that participants' perceived knowledge of the robot increased with additional exposures over time; however participant responses to open-ended questions about the robot's behavior and functions remained divergent over multiple scenarios. A discussion of the implications of apparent differences in human interpretation and prediction of robotic behavior and functionality is presented.

**Keywords:** human-robot interaction, mental models, perception of behavior.

## 1 Introduction

Advances in technology will enable robotic systems with greater intelligence and autonomy. In order for robots and other intelligent systems to succeed as useful teammates for humans, it is necessary to examine the impact of increased decision making capability of robots on individuals that interact with these systems [1]. Humans currently interact with robots in civilian and military contexts. However, in the current settings, the human largely decides the actions that are to be taken by the robot and initiates their execution via teleoperation [2]. In contrast, when robots are able to select and execute actions on their own, the burden falls upon the human to understand and interpret the behaviors and intention of robots. Human understanding of robots is, and will continue to be, further obscured by issues of robot reliability and human perceptions of trust, respectively [3]. In this document, we illustrate the influence of human understanding of robots within a Human-Robot Interaction (HRI) scenario, using results from an exploratory laboratory study in which novice users were tasked with observing, interpreting, and predicting robotic behavior.

While technology aspires to endow robots with mental models, there will also be an increased need for humans to hold an accurate mental model of the robot's mental model (i.e., understanding of the information/means by which robotic systems arrive at decisions and carry out actions). Mental models are the knowledge structures by

which humans organize information, and provide the mechanisms by which humans can explain system behavior and intention, as well as anticipate future behavior [4]. Variation in the scope and level of detail of mental models occur at the individual level, and the usefulness of a mental model depends on the manner in which it is applied to the situation at-hand. [5]. Therefore, mental models drive HRI [6]. The problem is that the level of sophistication of robot intelligence will likely be irrelevant if a human cannot understand how, what, or why the robot is acting in a certain manner.

Correct, accurate mental models of robots can be actively cultivated, for example, through design and training [7]. However, training is just one way to support accurate user mental models of robots, and it is not a panacea for haphazard design. Instead, multiple stakeholders must participate in the development of a user's mental model [8]. Whether intentional or unplanned, engineers, for example, specifically influence a user's mental model of a robot through their choices regarding physical characteristics [9], communication aspects [10], and robotic movement [11]. However, individuals perceive and react to robots in unique ways [12]. Additionally, humans have a propensity to apply social stereotypes to technological systems [13]. Therefore, the same robot's decision and action may be interpreted differently across users.

Similarly, mental models are naturally evolving systems that develop and change with experience. Through interaction, an individual will continuously modify his or her mental model in order to achieve an effective outcome [8]. This will be of particular importance for cultivating accurate mental models of robots for novice users, without necessitating extensive training. Since novice mental models of robots tend to be inaccurate and overly presumptuous [10], opportunities for interaction and acclimation will be important for fostering mental models that are true and correct representations of system capabilities and limitations.

The purpose of the investigation reported in this paper was to gain a better understanding of the scope and type of knowledge structures that humans, who had limited HRI experience, hold of robotic teammates; and to study the degree to which these knowledge structures can change with exposure to a robotic teammate. In addition, we examined differences in human interpretation of robotic behavior and intention. The results of this study highlight important considerations for the development of decision making capabilities of robotic teammates, especially in terms of designing for ease of human understanding of robotic behavior.

## 2 Method

This study was part of a larger data collection effort that included an investigation of mental model priming, previously held attitudes towards robots, and different techniques through which mental models might be assessed. For this paper, we specifically sought to examine the following hypothesis:

- *Hypothesis*: Over time, as participants are exposed and acclimated to their robotic partner, they report higher scores on a self-reported perceived mental model measure that pertains to knowledge of their team (self and robot), their task, their equipment, and the interaction between members of their team.

## 2.1 Measures

**Mental Model Survey.** This survey contained a series of questions regarding the degree to which participants had perceived knowledge of the task, team, team interaction, and equipment that was shared between the participant and a robotic entity. These questions were generated to be representative of the four types of mental models shared in teams as proposed by Mathieu and colleagues [14]. As such, the Mental Model Survey contained four subscales which included perceived knowledge of task, perceived knowledge of team, perceived knowledge of team interaction, and perceived knowledge of equipment. Participants utilized a 7-point Likert-type scale to indicate their responses to each item. Higher scores represented higher self-assessments of perceived knowledge of the item in question. Example items include, “The robot has knowledge of likely outcomes of this task” and “I understand this technology.”

**Free-Response Items.** Participants were asked to respond to several free-response items intended to gain a better understanding of differences in mental model interpretation of a simulated robotic teammate. Free-response questions were intended to probe participant mental models for their ability to explain system behavior, describe system functioning, and predict system actions. Example free-response items included “How would the robot signal to the Soldier that it has spotted something?,” “If the robot did find something, what action(s) would it take next?,” “What was the meaning of the gesture made by the robot at the end of the video?,” “How would the robot signal that it has spotted something?,” and “What would the robot do next?”

## 2.2 Participants

Fifty-one undergraduate students from a large southeastern university participated in this study. Participants’ ages ranged from 18-31 years ( $M = 20.09$  years,  $SD = 2.88$ ). Participants were recruited through the university’s research participation system and were offered credit in return for their participation.

## 2.3 Simulation Environment

Participants observed a series of video clips captured from the RIVET (Robotic Interactive Visualization & Exploitation Technology) computerized simulation, developed by General Dynamics Robotic Systems (GDRS). RIVET was built upon the Torque game engine, features a world editor, and contains a number of pre-configured environments, character models, and vehicles. The RIVET software allows multiple players to enter into a virtual environment and operate a variety of simulated unmanned ground and aquatic platforms. Additionally, the simulation environment can be networked with hardware in the loop (HITL) to evaluate sensor algorithms and software code.

Videos were created using a two-player, man-behind-the-curtain configuration. It was important to capture video clips from the perspective of an observer, rather than from the viewpoint of the robot. The player-observer viewpoint was intended to



simulate that of a Soldier working with an autonomous robot in an urban environment. Another player controlled the actions of a robot (a Talon-type robot, in this investigation) who executed a series of navigation and inspection and reconnaissance tasks, either alone or in the presence of civilians. Video was captured using a COTS product, Fraps, which captures video from a specified computer desktop window. The average length of each video clip was one minute.

## **2.4 Procedure**

After reviewing informed consent documents, participants completed a demographic questionnaire that contained general biographical information including familiarity with and attitudes towards computers, robots, and video games. However, these measures were a part of another study component that is not reported here.

After these preliminary activities, participants were provided a training presentation that presented the over-arching narrative for the videos they would be viewing, and familiarized the participant with the robot they were about to see. After viewing the training, participants completed the Mental Model Survey.

Participants then watched a series of 12 videos, broken up into two blocks. Each video depicted a human-robot team in which the robot was autonomously performing a series of inspection or reconnaissance-like tasks in an urban environment. For example, one video clip was presented with the introduction, “the robot was instructed to search the surrounding area for explosive materials.” For each video, the actions of the robot were congruent with the narrative (i.e., it did not inexplicably crash into a wall). Participants, however, were asked to interpret the individual motions and gestures made by the robot in the free-response survey.

Block order was counter-balanced between participants; videos within each block were randomized across all participants. The videos in each block differed in whether or not civilians were present while the robot was working. For each of the 12 videos (six in each block), participants were asked to carefully pay attention to the video, and then completed three open-ended, short-answer questions concerning their understanding of the robot, including what it was doing, (functionally) how it completed the work, and what it might do next. A smaller subset of these questions consisted of situational awareness items, designed to identify a participant’s engagement (or lack thereof). After each block, participants again completed the Mental Model Survey. Finally, participants were debriefed and thanked for their time.

## **3 Results**

### **3.1 Mental Model Development**

A one-way, repeated measures ANOVA was conducted to compare self-reported perceived knowledge of their robotic partner across three periods of time. The mental model measure was administered: immediately following training, after viewing the first block of videos depicting the robotic teammate, and after viewing the second block of videos depicting the robotic teammate. There was a significant main effect

for time, Wilks' Lambda = 0.696,  $F(1, 51) = 10.936$ ,  $p < .0005$ , multivariate partial eta squared = .304. A post hoc analysis with Bonferroni correction was conducted to determine whether a significant difference in reported knowledge of the robotic teammate was present each time perceived knowledge was measured, or if significant differences were present only at specific measurement opportunities. The test for simple effects revealed that there was a statistically significant increase in self-reported, perceived knowledge of the robotic teammate after viewing the second block of videos depicting the robotic teammate in the task environment ( $M = 21.67$ ,  $SD = 3.52$ ),  $t(51) = 4.37$ ,  $p < .0005$ . In addition, there was a statistically significant increase in self-reported, perceived knowledge of the robotic teammate between when the participants received the training ( $M = 20.42$ ,  $SD = 3.08$ ) and after viewing the second block of videos depicting the robotic teammate in the task environment ( $M = 21.67$ ,  $SD = 3.52$ ),  $t(51) = 3.20$ ,  $p = .002$  (see Table 1).

**Table 1.** Table of post hoc contrasts, means, and standard deviations

Contrast	Repeated Measure	Mean	Standard Deviation	Sig.
Pair 1	MMS post training	20.42	3.08	.703
	MMS post video block 1	20.57	3.12	
Pair 2	MMS post video block 1	20.57	3.12	.000*
	MMS post video block 2	21.67	3.52	
Pair 3	MMS post training	20.42	3.08	.002*
	MMS post video block 2	21.67	3.52	

Note: MMS = Mental Model Survey.

\*Statistically significant at  $p < .017$ . P value adjusted for Bonferroni correction.

### 3.2 Interpretation and Prediction of Robot Behavior

A qualitative analysis was conducted to examine apparent differences in perceived knowledge of a robotic teammate as indicated by the free-response items. An independent rater was used to look for thematic similarities across all participants for each of the free response items. Results could not confirm a specific hypothesis. However, the analysis did reveal that participants had thematically different understanding of robotic behavior. For example, in response to the item that asks, "What was the meaning of the gesture the robot made at the end of the video?," 14 of the participants made a thematic response that the gesture was intended to indicate that the area being searched was safe/no dangerous material had been found. Ten of the participants had a thematic response that indicated the opposite; the gesture made by the robot was an indication to the Soldier that the area was unsafe/hazardous materials had been found. Nine participants provided responses that indicated that they were unsure of meaning of the gesture or that the gesture was related to indicating something else like finding an object (without specific reference to safety) or the robot had finished its task. A sample of additional questions and corresponding thematic results are reported in Table 2.

**Table 2.** Sample of interpretations and predictions of the robot’s behavior and functions

Video description	Free-response question	
	Thematic category	Characteristic responses
A suspicious object was reported in an alley near the market place. The robot was instructed to locate the item and report back to the Soldier.	<i>How would the robot signal to the Soldier that it has spotted something?</i>	
	Arm movement	”Robot will raise arm.”
	Light	”Robot will turn on light indicating something has been found.”
	Sound	”Robot will produce a noise.”
	Electronic report	”Robot will message through radio.”
The robot was instructed to perform a routine inspection of the marketplace booths before civilians arrive, then report back to the Soldier.	<i>What was the meaning of the gesture* the robot made at the end of the video?(*Note to reader: up and down movement of manipulator arm)</i>	
	Safe	”Robot was nodding that the way was clear.”
	Unsafe	”Robot was nodding that something dangerous had been found.”
	Neutral (Item found)	”It was reporting that something was identified.”
	Unsure	”I do not know what the gesture meant.”
The robot is instructed to perform a thorough investigation of a burnt out car for traces of explosive material to determine the cause of the damage.	<i>Describe the equipment the robot uses to detect explosive materials.</i>	
	Arm	”Robot can use arm to touch and maneuver objects.”
	Camera on arm	”Robot has a camera on the top of its arm to look for explosives.”
	Sensor equipment	”The robot has a chemical scanner on its top of its body.”
	Unsure	”I am unsure how the robot would do this.”
The robot is tasked with navigating to and inspecting barrels for explosive materials in an alley of a small village. After inspecting the barrels the Robot must return to the Soldier and await further orders.	<i>At what distance is the robot able to detect explosive materials?</i>	
	Close	”Probably close, like 3-5 feet away.”
	Far	”A far enough distance to be away from a blast.”
	Unsure	”I don’t really know.”

## 4 Discussion

The hypothesis was partially supported, as results revealed a significant main effect for time on change in reported knowledge of the robotic teammate. An examination of the simple effects revealed that there was a significant difference in reported mental models between time 1 (immediately post training) and time 3 (following viewing all of the videos). In addition, there was a significant difference in reported knowledge of the robotic teammate between time 2 (after viewing the first block of videos) and time 3 (after viewing the second block of videos). This finding lends support for the importance of experience when forming mental models of robotic teammates.

For example, while participants' reported that their knowledge of the robotic teammate steadily increased over time, the difference between their perceived knowledge post-training (i.e., but before viewing any videos) and after viewing one block of videos was small and non-significant (see Table 1). That is, there was no significant effect for the order in which the two blocks of views were presented. Only after viewing the robotic teammate in both blocks of videos, which presented the teammate performing tasks in two different contexts (i.e., civilians present while the team was working and civilians not present while the robot was working), was there a significant difference in self-reported knowledge of the robotic teammate. This may indicate that familiarizing the human partner to the robot in multiple contexts is important to developing a clear understanding to the robot.

The data suggests that subjective self-assessment of participant mental models increase over time, as participants see the robot performing a wide variety of tasks in different contexts. However, this did not imply that mental models between participants were similar. For example, when asked how a Talon robot might inform a Soldier that it encountered an IED, participant responses varied greatly. Some participants applied anthropomorphic stereotypes and social rules to the robot (e.g., it should wave its arm; it should nod). Differences in the expected method of communication were also observed across participants (e.g., it would produce a noise to inform the Soldier it found something). These expectations were also used to reason about the robot's performance on the task (e.g., I am assuming it checked everything along the path it was instructed to check; the open garage was obvious, it should have looked there), as well as how the robot would act in the future.

## 5 Conclusion

This research highlights a number of important considerations for the development of high-autonomy, intelligent robots. Specifically, an individual's perceived understanding of a robotic system increases given additional robotic exposures in different contexts. However, it is important to recognize the difference between the richness of understanding and accuracy of understanding. Different individuals may be similarly confident in their knowledge of the robot, but arrive at vastly different conclusions about its function and performance; with each conclusion being an equally valid

interpretation of the robot's behavior (e.g., "it signaled that it detected something" vs. "it signaled that it did not detect anything").

Secondly, given a task-oriented human-robot team, the ability of an autonomous robot to complete the task is only part of the overall solution. It is important to consider additional functions such as: How will the robot indicate the task has been completed? How will the user know if the robot cannot complete the task? What should the robot do if it becomes stuck or damaged? What information should the human expect to provide to the robot? And, what information should the human expect to receive from the robot?

Finally, it is safe to assume that individuals, specifically designated to be robot handlers, will receive adequate training to mitigate the effect of potentially ambiguous robot behavior. However, robots will be expected to operate among other team members, bystanders, and even hostile forces for which specialized training will be minimal or non-existent. It is therefore necessary to consider those cases in which the intention of robot behavior should be transparent, opaque, or even deceptive, depending upon the circumstances of the operational environment.

We continue to investigate the role of human understanding within HRI, in order to provide designers useful input for the development of robotic systems that are compatible with the knowledge and understanding of their human counterparts.

**Acknowledgements.** The research reported in this document/presentation was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory. The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

1. Phillips, E., Ososky, S., Grove, J., Jentsch, F.: From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, pp. 1491-1495 (2011); Murphy, R.R.: Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 34(2), 138-153 (2004)
2. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53, 517-527 (2011)
3. Rouse, W.B., Morris, N.M.: On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin* 100, 349-363 (1986)
4. Johnson-Laird, P.N.: *Mental models: Toward a cognitive science of language, inference, and consciousness*. Harvard University Press, Cambridge (1983)

5. Lohse, M.: Bridging the gap between users' expectations and system evaluations. In: 2011 IEEE RO-MAN, pp. 485–490
6. Wickens, C.D., Hollands, J.G.: Engineering psychology and human performance. Prentice-Hall, Upper Saddle River (1999)
7. Norman, D.A.: Some observations on mental models. In: Gentner, D., Stevens, A.L. (eds.) *Mental Models*, pp. 7–14. Lawrence Erlbaum Associates, Inc., Hillsdale (1983)
8. Sims, V.K., Chin, M.G., Sushil, D.J., Barber, D.J., Ballion, T., Clark, B.R., Garfield, K.A., Dolezal, M.J., Shumaker, R., Finkelstein, N.: Anthropomorphism of robotic forms: A response to affordances. In: *Human Factors and Ergonomics Society Annual Meeting*, Orlando, FL, vol. 49, pp. 602–605 (2005)
9. Lee, S., Kiesler, S., Lau, I., Chiu, C.: Human mental models of humanoid robots. In: *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 2767–2772
10. Ju, W., Takayama, L.: Approachability: How people interpret automatic door movement as gesture. *International Journal of Design* (2009)
11. Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics* 24, 442–451 (2008)
12. Nass, C., Moon, Y.: Machines and mindlessness - social responses to computers. *Journal of Social Issues* 60, 81–103 (2000)
13. Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: Influence of shared mental models on team process and performance. *Journal of Applied Psychology* 85, 273–283 (2000)

# Computational Mechanisms for Mental Models in Human-Robot Interaction

Matthias Scheutz

Department of Computer Science  
Tufts University, Medford, MA 02155, USA  
[matthias.scheutz@tufts.edu](mailto:matthias.scheutz@tufts.edu)  
<http://hrilab.tufts.edu/>

**Abstract.** Mental models play an important and sometimes critical role in human-human interactions, in particular, in the context of human team tasks where humans need to interact with each other to achieve common goals. In this paper, we will describe some of the challenges involved in developing general computational mechanisms for mental models and their applications in the context human-robot interactions in mixed initiative tasks.

## 1 Introduction

The rapid advances in robot technology over the last two decades has enabled a shift in robot applications from very confined and constrained industrial settings to more open unconstrained human-like environments (from hospitals and elder-care settings, to offices and households). Increasingly, robots are also envisioned to become part of “mixed-initiative” teams, where humans and robot need to collaborate to achieve common goals. A case in point is the initial 2011 “National Robotics Initiative” funding program of the National Science Foundation in the US which explicitly targets “innovative robotics research and applications emphasizing the realization of [...] co-robots acting in direct support of and in a symbiotic relationship with human partners”, where “co-robot” is a term specifically coined to denote robots that “work beside, or cooperatively with, people”.<sup>1</sup>

Common to the wide range of co-robot applications (from search and rescue missions in disaster zones, to space exploration scenarios) is the role of the robot as a genuine helper, a true team member that acts in the interest of the team in a reliable and effective manner, just as a human team member would. Obviously, this is a very high bar for robots to meet, for many reasons. First of all, the robot has to be able to perform the task-based activities for its envisioned role. For example, a search and rescue robot might have to be able to perform triage on a wounded person or at least be able to instruct another human on how to do it [7]. This, itself, could be very challenging, e.g., if complex manipulation capabilities are required such as administering a syringe as part of the first-aid procedure or

---

<sup>1</sup> See <http://www.nsf.gov/pubs/2011/nsf11553/nsf11553.htm>

repairing broken pipes and valves in a power plant accident. Moreover, the robot has to be able to function reliably and autonomously for possible long periods of time without human intervention (e.g., in an underground sewer system, the rubble of collapsed buildings, or an extra-terrestrial space station setting with limited to no network connectivity).

While both of the above challenges will require sustained research efforts for years to come, there is at least one other challenge that has not received sufficient attention yet, even though it is potentially even more difficult to address than the two engineering challenges: that of *effective, natural human-robot interaction* (HRI) [11]. “Natural HRI” here means that humans will be able to communicate and interact with robots in the mixed initiative settings *as if* those robots were humans. For example, in a search and rescue scenario, a human commander might have to interact in natural language with a robot remotely located in a collapsed building about where to search for victims. And even though the dialogue exchanges will be focused on the task at hand, the underlying architectural and computational requirements on the robotic side for enabling even simple task-based natural dialogues are quite astounding [10].

Of the many open research problems in natural human-robot interaction, we will focus on a critical underlying capability that affects almost every aspect of the robotic control system: *the robot’s ability to build and maintain mental models of other (human and robotic) team members*. We start by briefly reviewing the context of mental models in human teams and then describe the challenges involved in developing computational mechanisms for robotic control architectures for mixed initiative scenarios. Specifically, we will point to several different kinds of representations that are needed for building sufficiently accurate mental models of team members that, in turn, allow for making predictions and coordinating, at least up to some level, joint activities.

## 2 Motivation: Team Mental Models

It is well-known from extensive research in human teaming that for team members to coordinate their activities effectively and achieve overall high performance, they need to keep track of each other’s mental states such as goals and subgoals, intentions, beliefs, and various others – a process commonly subsumed under the term “shared mental models” (e.g., [6]). “Shared mental models” are related to and could be subsumed under the concept of “mental model” as used in psychology, where it typically refers to the types of hypothesized knowledge structures humans build in order to make sense of their world, to make inference based on the available information and to make predictions about future states (e.g., [8]). However, while mental models research in psychology has more focused using mental models to explain various types of human reasoning, mental models in the context of teams have more to do with establishing and maintaining *common ground* (e.g., in Clark’s sense [4]) and building *team mental models* [3] that aid in decision-making and the adjustment of one’s behavior based on predictions made about the other team members’ future activities and actions.



Thus, team mental models are critical for making sense of team activities, for understanding the dynamic changes of team goals and team needs, the possibly dynamic changes of roles and functions of team members, and the overall state of the team.

While there is increasing evidence, especially from research in management and organizational behavior, that humans build and use team mental models, and that using them appropriately will lead to improved team performance, little is known about what these mental models look like, i.e., what information is represented about other team members, their intentions and goals, their knowledge and beliefs, and their activities. Moreover, it is not clear exactly how these representations influence the various cognitive processes involved in coordinating team activities, from task-based natural language dialogues, to distributed task and role assignments, to team decision-making. Yet, these aspects are critical for understanding the functional mechanisms of team mental models at a level that would allow for the development of similar mechanisms in robots. Note, however, an important difference in the aim of implementing team mental models in robots compared to providing computational models of human team mental models, say: since robotic and human cognitive architectures are (currently) very different, we do not claim nor intend to provide a computational account of how humans build and use team mental models; rather, we will attempt to lay out some of the important data representations and processes that have the potential to improve robot operation and might make robots better team mates.

### 3 How to Formulate and Represent Mental Models in Robots

Mental models, to be usable in robotic architectures, effectively need to be broken down into two parts: the *data representations* that capture information about the task, the other team members, and the environment, and the *computational processes* that operate on these data structures to create, maintain, revise and discard them. The former will likely differ from task to task, while the latter are intended to be more general mechanisms that can be used across tasks. To be able to motivate the required data structures and processes better, we will use a team search task from our prior work [5].

The specific team search task requires at least two humans, a remotely located commander and at least one searcher located in the target environment that is to be searched. All team members must coordinate their activities through spoken natural language interactions via wireless audio links as their only interaction modality (given that they are spatially separated). The commander has a rough map of the target environment, while the searcher does not have any map. Neither commander nor search have been in the environment before (in our experiments, the indoor search environment consisted of several rooms and a surrounding hallway). The team has two main tasks to begin with (later a third task was added): (1) the searcher has to inform the commander of any encountered green boxes as the commander has to mark their locations on the

map; (2) the commander has to direct the searcher through the environment to a particular location where the searcher can pick up a container which then can be used to collect colored blocks from blue boxes located in the environment.

The following example is one of many from our CReST corpus [5] showing how commander and searcher collaborate via natural language dialogues to agree on actions in the interest of their task goals:

Commander: Okay. Go through the open door and towards the steps that are right in front of you before the steps take a right.  
 Searcher: Okay, right. Right on the steps there's a green box number two.  
 Commander: Oh, number two right on the steps.  
 Searcher: Yeah.  
 Commander: Okay, I got it. Alright. If you're looking at the steps you take a right there should be another open door.  
 Searcher: So, don't actually go up the steps?  
 Commander: Don't actually go up the steps.  
 Searcher: Okay. Yep, I see the door.

Notice how the commander instructs the searcher where to go by describing the environment from the searcher's perspective based on a mental model of where the searcher is in the environment. These instructions frequently comprise descriptions of salient aspects of the environment as gleaned from the map such as "go through the open door and towards the steps that are right in front of you". Here, the director's description relies on a mental model of where the searcher is located and what the environment would look like from the searcher's perspective.

Also notice how the searcher interrupts the activity when she notices the green box on the stairs and communicates that to the director. This requires both parties to keep track of their overall activities as subactivities related to some of the task goals are initiated. After the commander confirms by repeating back the number and location of the box, the search confirms, and the subactivity finishes. This requires both parties to resume the previous activity of the commander directing the searcher through the environment. Here it is interesting to note that the searcher had a different goal in mind (namely to go up the stairs) from what the commander intended (to turn right), and the instruction to "take a right" by the commander revealed the goal discrepancy to the searcher, thus prompting the searcher to ask explicitly about the next action ("don't actually go up the steps?"), which in turn revealed her goal to the commander. Again, this interruption constitutes a sub-task of achieving goal alignment and consolidating the mental models of both parties about what the next actions and subgoals are. Once, agreement is reached, the searcher informs the commander of the fact that she can see the door, which directly answers the previous "hedged explain" dialogue move by the commander that "there should be another open door".

These types of dialogue moves specifically require the interactant to take verification action (i.e., to verify that there is another open door) and confirmation or disconfirmation action (i.e., that the open door was verified) [5].

All of the above dialogue interactions serve the dual purpose of establishing what actions to take next (i.e., common subgoals in the interest of the overall task goals) and to establish *common ground* between searcher and commander [4]. Common ground here comprises several aspects such as where the searcher is located, what the searcher's perspective is, what objects the searcher can see, what goals the searcher/commander has, etc. Common ground is negotiated through dialogue interactions which eventually are finished with acknowledgments of both parties and various dialogue-based linguistic mechanisms are employed to communicate understanding or lack of confidence.

Aside from important dialogue-based mechanisms for negotiating goals and activities which are interesting in their own right, the above scenario points to several important aspects that mental models need to capture: (1) *facts*, about the task, events, objects, and the environment, including aspects about the location of team members; and (2) *beliefs*, about goals, activities, and beliefs about team members. The difference between facts and beliefs here is that facts are taken to be true from an agent's perspective while the content of beliefs might not be true. For example, the searcher believed that the goal was for her to go up the steps, while the commander intended for the searcher to turn right. For her to be able to detect the belief inconsistency, she had to represent this goal and keep it in her mental model of the commander. When the commander then gave an instruction that suggested another incompatible goal, the inconsistency was discovered. Hence, it is critical to represent the other agents' perspectives, which will allow for better detection of inconsistencies and thus improved alignment among agents.

In addition to the rich data representations, it is important for an agent to allow for belief maintenance and belief revision processes that can both synchronize beliefs and update them when new evidence arrives. This does not only apply to the beliefs of other agents as represented in the agent's mental models, but also to the facts the agent holds true (for it is possible that these factual representations were obtained from insufficient or flawed evidence, misinterpreted communications, etc.). Moreover, an agent not only has to correct her own inconsistent facts and beliefs, but those corrections might, in turn, trigger corrections of other agents' facts and beliefs (e.g., if the agent learned that a previously communicated fact is not true).

## 4 How to Build and Update Mental Models

The above interactions clearly showed that representations used to build mental models need to be sufficiently expressive to be able to capture goals, beliefs, desires, and knowledge-based states (such as rules and facts), as well as various modal operators (e.g., about beliefs of other agents and their beliefs about other agents, but also possibilities, obligations, permissions, etc. of actions, but also).

Here, we will build on our previous pragmatic and mental modeling framework [1,2] to be able to discuss some of updating processes needed for mental models. To simplify the discussion, we will use  $[[\cdot]]_c$  to denote the “pragmatic meaning” of an expression (e.g., a natural language utterance) in some context  $c$ , which in team tasks typically includes task and goal information, as well as belief and discourse aspects. And we will use  $\alpha$  to denote the agent under consideration (i.e., the agent whose whose perspective we will take for the discussion of how to update the agent’s mental model).

Overall, updates to an agent’s mental model will be triggered by various events, mediated through the agent’s perceptual system. For example, the agent might perceive a new task-relevant object (as in the case of the searcher above noticing a green box). Assuming that the agent will store this perception, we can formulate a general principle that if  $\alpha$  perceives an object  $o$  at location  $l$  at time  $t$ , then  $\alpha$  will believe (B) that it perceived  $o$  at  $l$  at  $t$ :

$$\text{Perceives}(\alpha, o, l, t) \Rightarrow B(\alpha, \text{Perceives}(\alpha, o, l, t))$$

This principle can then be applied to all agents on the team. As a special case, if  $\alpha$  perceived another agent  $\beta$ , it will form the belief that it perceived  $\beta$  and it might also form the belief that  $\beta$  perceived  $\alpha$ . Similar principles can be defined for actions to capture their enabling conditions and effects (i.e., pre- and post-conditions).

More interesting are belief updates that have to do with perceptions that are communications, i.e., when  $\alpha$  receives an update from  $\beta$  through an utterance  $Utt$  in a given context  $c$ . Given that all team members are collaborating on a common set of task goals and thus have no incentive to purposefully mislead or deceive their team members, we can assume that all communicated propositions  $\phi \in [[Utt]]_c$  are true (at least from  $\beta$ ’s perspective). It is then necessary for  $\alpha$  to check whether any of the communicated propositions are inconsistent with  $\alpha$ ’s existing beliefs. For this purpose,  $\alpha$  needs to employ an inference algorithm  $\Rightarrow_\alpha^b$  that will, up to some bound  $b$ , check for inconsistencies.<sup>2</sup> Any inconsistent conflicting beliefs are then removed from the agent’s sets of beliefs  $Bel_\alpha$  (e.g., in the above example, the searcher would remove the goal to go the stairs as a result of the commander’s correction telling the searcher to go right instead).

Moreover,  $\alpha$  believes all propositions it can infer from (again within bound  $b$ ) from propositions  $\phi \in [[Utt]]_c$ :

$$([[u]]_c \Rightarrow_\alpha^b \phi) \wedge \text{Heard}(\alpha, u) \Rightarrow B(\alpha, \phi)$$

By the same token,  $\alpha$  also believes everything it says:

$$([[u]]_c \Rightarrow_\alpha^b \phi) \wedge \text{Said}(\alpha, u) \Rightarrow B(\alpha, \phi)$$

---

<sup>2</sup> Note that the bound  $b$  is intended to capture both the agent’s reasoning limitations as well as other time-based limitations based on the current context.

In addition to updating its own beliefs,  $\alpha$  needs to model any other agent  $\gamma$  also hearing  $\beta$ 's utterance, i.e.,  $\alpha$  has to derive its mental model  $\{\psi | B(\gamma, \psi) \in Bel_\alpha\}$  for all other agents  $\gamma \neq \alpha$  and update it by using the same rules it applies to its own beliefs. The same is true if  $\alpha$  notices that another agent has certain perceptions or performs certain actions. In general,  $\alpha$  needs to update all its models whenever there is a change in its own beliefs (either through addition of a new fact or revision of a known one).

In team tasks, the situation often arises that another agent “intends to know” (IK) a proposition, i.e., either makes an explicit request to be informed or provides some other information from which such a request can be derived. The set of all propositions other agents want to know,  $\Phi_{IK}$ , can be defined as:

$$\psi \in \Phi_{IK} \Leftrightarrow \exists \beta, \phi : \psi \in Bel_\alpha \wedge IK(\beta, \phi \in Bel_{alpha}) \wedge (\psi \Rightarrow_\alpha^b \phi \vee \psi \Rightarrow_\alpha^b \neg\phi)$$

This set of proposition is then part of what agent  $\alpha$  needs to communicate to other agents, in addition to the set of all propositions that will correct false beliefs  $\Phi_{rev}$  that agents holds, defined as:

$$\psi \in \Phi_{rev} \Leftrightarrow \exists \beta, \phi : B(\beta, \phi) \wedge \phi \in Bel_\alpha \wedge (\psi \Rightarrow_\alpha^b \neg\phi)$$

The above principles have been integrated in a cognitive robotic architecture for human-robot interaction and tested in simple human-robot interactions [1,2]. However, a thorough experimental evaluation with naive subjects has not yet been performed.

## 5 Discussion

The previous sections briefly sketched the kinds of principles necessary for building mental models in computational architectures, in particular, what representations to build and maintain, how to update them based on events and internal changes (e.g., results of inferences), and when to communicate them to provide answers to requests or correct false beliefs of other agents. While the principles were stated in fairly general ways, there are important aspects of their computational implementation that were only hinted at. For example, computing deductive closures of any finite set of beliefs is usually not feasible for reasonable inference rules, hence a “bound” was imposed on the inference procedure. However, even such a bound might still make various computations intractable. E.g., if there is a large number of agents with many different belief and knowledge items, then even keeping track of all of them and detecting inconsistencies among them without any inference might not be feasible in a reasonable amount of time. On the other hand, it is doubtful that humans could do that either. Hence, restricting the above principles to small, human-like teams might be sufficient for addressing the computational overhead.

A related question that can also lead to computational escalation is the level of nested beliefs an agent should keep track of (e.g., it is necessary to keep track of beliefs of beliefs of beliefs?). The answer here is largely pragmatic, for in some

cases this type of information may not even be available or if it is, it might be too short-lived to be of interest (e.g., if a third agent observed the dialogue in the human search task between the search and the commander about whether the searcher should go up the stairs, it could have represented the goals of the searcher and the commander before, during, and after the dialogue, and updated them throughout as it became clear that the searcher had to revise its goal; but none of that processing might have impacted the activity of this third agent). Hence, there is a critical notion of “relevance” to one’s own activity as part of the team that should be taken into account when building and updating mental models. What this notion is and how it can be cast in algorithmic terms is, however, an open question at this point.

Other interesting questions are related to the extent to which multiple agents will manage to keep their mental models synchronized depending on their communication abilities and the task demands. It might be possible to impose hierarchical structures that would allow agents to focus on a subset of the team members and only track their activities and goals. Such hierarchical structures could also help curb the computational overhead of mental model processing.

Finally, it is also important to consider ways to evaluate the efficacy and efficiency of mental models employed on robots. Here evaluation measures that have been used in human teams might be applicable (e.g., [9]) to the extent that the measures are objective and can be answered directly from observations. Subjective measures (such as questions asked of human team members after the tasks) could also be employed as long as they can be obtained from logged information about architecture-internal states (e.g., the set of beliefs about other agents beliefs could be examined with respect to the accuracy of those beliefs throughout the task).

## 6 Conclusions

In this paper, we discussed some of the computational mechanisms required for building and maintaining mental models of team members in mixed-initiative human-robot teams. We started by examining typically task-based natural language interactions in human teams and derived some representational and processing requirements from those interactions. We then introduced a formal framework for representing and updating mental models in a way that they can be integrated into a robotic architecture. Finally, we discussed some of the challenges involved in employing such mental models given the computational and real-time constraints of robots working with humans and pointed to some interesting future directions that will require extensive experimentation and modeling of human-robot teams in order to determine the extent to which mental models need to be built and maintained in order to improve team performance.

## References

1. Briggs, G., Scheutz, M.: Multi-modal Belief Updates in Multi-Robot Human-Robot Dialogue Interactions. In: Proceedings of AISB 2012 (2012)
2. Briggs, G., Scheutz, M.: Facilitating mental Modeling in Collaborative Human-Robot Interaction through Adverbial Cues. In: Proceedings of 12th Annual SIG-DIAL Meeting on Discourse and Dialogue, pp. 239–247 (2011)
3. Cannon-Bowers, J.A., Salas, E.: Cognitive psychology and team training: Shared mental models in complex systems. Paper Presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, Miami, FL (1990)
4. Clark, H.H.: Using language. Cambridge University Press, Cambridge (1996)
5. Eberhard, K., Nicholson, H., Kübler, S., Gundersen, S., Scheutz, M.: The Indiana 'Cooperative Remote Search Task' (CReST) Corpus. In: Proceedings of Language Resources, Technologies and Evaluation LREC 2010 (2010)
6. Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., Cannon-Bowers, J.A.: The Influence of Shared Mental Models on Team Process and Performance. *Journal of Applied Psychology* 85(2), 273–283 (2000)
7. Harriott, C.E., Zhang, T., Adams, J.A.: Evaluating the applicability of current models of workload to peer-based human-robot teams. In: Proceedings of the 6th International Conference on Human-robot Interaction, pp. 45–52 (2011)
8. Johnson-Laird, P.N.: *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press (1983)
9. Lim, B.-C., Klein, J.K.: Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behaviour* 27, 403–418 (2006)
10. Scheutz, M., Cantrell, R., Schermerhorn, P.: Toward humanlike task-based dialogue processing for human robot interaction. *AI Magazine* 32(4), 77–84 (2011)
11. Scheutz, M., Schermerhorn, P., Kramer, J., Anderson, D.: First Steps toward Natural Human-Like HRI. *Autonomous Robots* 22(4), 411–423 (2007)

# Increasing Robot Autonomy Effectively Using the Science of Teams

David Schuster and Florian Jentsch

University of Central Florida, Institute for Simulation and Training  
3100 Technology Parkway, Orlando, FL 32826  
dschuster@ist.ucf.edu, florian.jentsch@ucf.edu

**Abstract.** Even as future robots grow in intelligence and autonomy, they may continue to face uncertainty in their decision making and sensing. A critical issue, then, is designing future robots so that humans can work with them collaboratively, thereby creating effective human-robot teams. Operators of robot systems can mitigate the problems of robot uncertainty by maintaining awareness of the relevant elements within the mission and their interrelationships, a cognitive state known as situation awareness (SA). However, as evidenced in other complex systems, such as aircraft, this is a difficult task for humans. In this paper, we consider how application of the science of human teaming, specifically task design and task interdependence in human teams, can be applied to human-robot teams and how it may improve human-robot interaction by maximizing situation awareness and performance of the human team member.

**Keywords:** Human-robot interaction, system design, situation awareness.

## 1 Introduction

Even as future robots grow in intelligence and autonomy, they may continue to face uncertainty in their decision making and sensing. A critical issue, then, is designing future robots so that humans can work with them collaboratively, thereby creating effective human-robot teams. Operators of robot systems can mitigate the problems of robot uncertainty by maintaining awareness of the relevant elements within the mission and their interrelationships, a cognitive state known as situation awareness (SA). However, as evidenced in other complex systems, such as aircraft, this is a difficult task for humans. Often, when automated systems fail, the result is a lack of situation awareness. This problem has come to be known as the out-of-the-loop performance problem [1].

To achieve high SA in the human-robot team, robots should supplement the knowledge held by their human team mates, and human team members, in turn, should be able to aid the robot(s) without sacrificing their own knowledge or experiencing high workload. We believe that the most efficient and effective way for designers to facilitate collaborative work with robots is by applying principles of human teamwork to human-robot systems.



While teaming can inform human-robot interaction, it is only one angle from which to approach this problem. In our other paper [2], we examine how robot design can mitigate or exacerbate the effects of the out-of-the-loop performance problem. In this paper, we specifically consider how the application of the science of human teaming, and, in particular, of task design for human teams, can improve human-robot interaction by maximizing situation awareness (SA) and performance of the human team members. Specifically, we propose that structuring a task to promote task interdependence between human and robot is a potential solution to improve SA in future human-robot teams.

## 2 Background

### 2.1 Situation Awareness and Human-Robot Interaction

Situation awareness (SA) describes the relevant knowledge held by an operator while performing a task. Endsley [3] created a model in which SA is a high-level, goal-directed information processing function as part of a sensation-decision-action cycle. Generally, SA is goal-directed, high-level knowledge that comes as a result of an individual's information processing within an environment [4].

In a team context, Salas and colleagues' suggested that SA is an individual-level process that takes place in the context of team process [5]. In human-robot interaction, this model is useful for explaining how heterogeneous agents (e.g., a robot and a human) may develop SA as part of a team. Robots differ from humans in how they acquire, process, and structure information. By investigating SA as an individual-level process, researchers and practitioners can focus on maintaining a sufficient level of SA in the human without asserting that the robot performs human information processing. This model also allows consideration of robots at any level of autonomy. Robots have been traditionally thought of as tools, and in this model, the SA of the human is independent of the information processing of the robot. The robot participates in the human's development of SA through its communication and coordination with humans. One way to assess the effectiveness of the human team member within the system is, therefore, to measure SA.

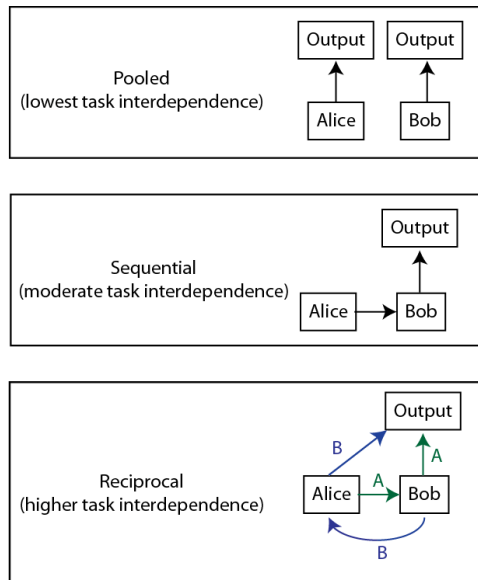
Researchers have aimed to increase SA when working with automated systems by strategically altering the level of automation. Two categories of solutions have come from this approach: intermediate levels of automation and adaptive automation. The former is characterized by utilizing a lower level of automation than may be technically feasible as a means of avoiding the effect [6]. Another solution is to employ adaptive automation [7]. Adaptive automation encompasses a number of methods aimed at optimally dividing tasks between humans or automated agents based on the state of the environment, system, or behavior of the human.

Both intermediate levels of automation and adaptive automation can effectively change the task assignments between automation (here: robot) and human, but they often do so at the task level: Task A is performed either by the human or the robot, and its assignment may be switched as the situation changes.

## 2.2 Task Interdependence

Recently, Johnson and colleagues [8] put forth a number of criticisms of solutions based on selecting the appropriate level of autonomy. Rather, they argued, “it is more productive to think about autonomy in terms of multiple task-specific dimensions rather than in terms of a single, uni-dimensional scale.” We aim to expand upon existing solutions to the creating and maintaining SA specific to robotic systems by including task/mission factors as a moderator of the relationship between autonomy and SA.

Task interdependence has been studied extensively as a performance factor in human teams. Task interdependence is the degree to which members must rely on each other to perform their tasks effectively given the design of their jobs [9]. An example of task interdependence is the necessity of sharing equipment or materials to achieve performance outcomes [10].



**Fig. 1.** Illustration of selected levels of Thompson’s (1967) framework of task interdependence in human teams

Thompson’s [11] categorization of task interdependence has been widely cited. In this classification, discreet interdependence styles are ordered from least to most interdependent (Fig. 1). At the lowest level of task interdependence, each agent performs equivalent subtasks. Each pooled subtask generates a complete output, and the agents are not at all interdependent. An example would be the workshop of multiple craftsmen in which each individual produces a complete product. The number of completed products is the measure of work output. Sequential task interdependence requires performance of a subtask by each agent in order to produce an output. A key quality of this level is that the order in which the tasks are performed is not flexible. In the figure, Alice must always perform her task before Bob can complete his.

Only when Bob completes his second task has the output been created. Reciprocal interdependence adds flexibility in the order but not the specialization of role. In the diagram, Alice and Bob each perform a necessary step, but those steps can be performed in any order (either path A or path B). An example of this kind of work in a complex system is a surgical team.

Thompson's classification takes an information-processing approach and characterizes organizations as information processing systems working under conditions of limited capacity [12]. Although this is not the only taxonomy of task interdependence, and task interdependence is not the only driver of interdependence within a human team [13], Thompson's taxonomy offers provides a clear operationalization of the construct, which is complementary to Endsley's information processing approach to SA. Although it has been studied primarily in human teams, it has been applied to automation as well.

**Task Interdependence and Human Teams.** Although task interdependence is an important construct in the team literature [14], there is limited support for a direct relationship between task interdependence and performance. Inconsistent findings of main effects suggest the presence of moderators of this relationship [15] [16] and the interaction of task interdependence with other determinants of performance [14] [17]. In a meta-analysis, LePine, Piccolo, Jackson, Mathieu, and Saul [18] found that task interdependence was a moderator of the relationship between teamwork (the coordinating behaviors amongst team members) and team effectiveness. Higher levels of task interdependence lead to stronger relationships between teamwork and effectiveness. In human teams, group control more positively impacted performance as task interdependence increased [19]. Although research on exclusively human teams does not have immediate implications for human-robot teams, it shows that task design interacts with other work factors in affecting the ability of humans to perform shared tasks.

**Task Interdependence and Automation.** As a task characteristic, task interdependence has been explored in human-automation interaction as well. Johnson and colleagues [20] suggested that lack of SA, such as in the out-of-the-loop performance problem, is the outcome of human-robot interaction under high autonomy, which they defined as a combination of self-directedness and self-sufficiency. They refer to the problem as opaqueness; the operator, disconnected from the behavior of the robot, cannot integrate or predict its actions, leading to reduced performance. Low levels of reliable autonomy lead to the robot being a burden on operator cognitive resources while the system becomes opaque at higher levels of reliable autonomy. Johnson et al. [20] presented the results of a study in which higher levels of autonomy, in the absence of task interdependence, lead to increased opaqueness and less subjective burden. In Johnson et al.'s work, task interdependence was described as an orthogonal third dimension that moderates this effect. Johnson et al. [20] argued that task allocation approaches do not allow agents to depend on each other and that increased levels of task interdependence are a means for operators to remain in the loop. By working in closer collaboration, operators may have more opportunities to communicate the

status of shared tasks while focusing on their individual level outputs. This effect has been observed in human teams [21]. However, task interdependence may only show a benefit when the robot can contribute by performing an independent subtask.

**Task Interdependence in Human-Robot Interaction.** Since the problems with maintaining SA in complex systems are not limited to robots, much of the work in this area has examined human-automation interaction and not human-robot interaction. Robots are a unique form of automated system. As with human teaming, general approaches to task allocation of automation must be extended to address aspects of the problem unique to human-robot interaction.

Robots provide the most benefit when they perform tasks that are unsafe or undesirable. In both cases, increasing task interdependence may appear to conflict with the desire to physically separate the human from the work performed by the robot. We offer an extension to the theory and findings of Johnson and colleagues [20] by suggesting that the information requirements of the task can be interdependent even if the physical work is not. Tasks that require human and robot to continuously and seamlessly share information may provide similar benefits as tasks that require a human to hand off physical work with a robot. In this way, the robot and human will adapt the mission-relevant knowledge to each other and keep the human in the loop.

Prior investigations of the out-of-the-loop performance problem have been limited to system factors and have not considered the impact of task/mission factors. Both have been shown to independently affect SA in human-robot interaction [22]. However, it is not known how task/mission factors and system factors may interact to affect SA. While modifications to the task may be appropriately disregarded when mission goals are fixed and the relationship between human and automation is well defined (for example, in a nuclear power plant), the capability of robots is continually changing. Thus, it is important to not only investigate when a robot should operate autonomously and how tasks should be designed, but also the kinds of tasks for which robot autonomy is best suited.

## 3 Conclusions

### 3.1 Situation Awareness as a Metric for Human-Robot Interaction

Our first proposition is that robot system designers can understand the knowledge held by the human operator and, consequently, improve system performance through measurement and maximization of SA. The robot's contribution to SA is through the provision of information needed by the human operator who, in turn, builds SA. Given its importance in current robots and other complex, automated systems [23], SA should be considered as a metric for operator knowledge within the system.

### 3.2 Task Interdependence as a Framework for Task Design

Our second proposition is that task interdependence moderates the relationship between autonomy and SA such that high levels of autonomy do not lead to poor SA

when task interdependence is high. That is, task interdependence is a potential solution to the out-of-the-loop performance problem. Further, the physical work of the task may not need to be interdependent if the knowledge needed to perform the task is interdependent, requiring collaboration between human and robot that does not conflict with individual-level goals. The success of human teams is impacted by task interdependence, although task interdependence does not directly affect performance. Rather, mediators suggest that task interdependence must be appropriate given other task factors.

In human-robot interaction, a critical need is to be aware of the robot while minimizing the time spent interacting with the robot. By employing robots in tasks that have reciprocal interdependence with the human team member, periodic sharing of information between human and robot team member takes place, and at the same time, advances mission goals. Consequently, the time spent interacting with the robot provides more efficient information sharing. An example of this would be a reconnaissance task in which the human and robot must monitor separate areas. This task becomes interdependent if individuals monitored by the robot affect or inform individuals monitored by the human. That is, monitoring separate rooms within a building could be reciprocal, whereas monitoring rooms in separate buildings would be pooled.

Our third proposition is an inverse of the first: Task interdependence moderates the relationship between autonomy and SA such that high levels of autonomy lead to poor SA when task interdependence is low. Tasks that are inadvertently designed with minimal task interdependence may exacerbate problems with maintaining SA. In this scenario, communication with the robot is not necessary for the accomplishment of mission goals, leading to a situation where the human team member must choose between accomplishing mission goals and interacting with the robot. Because interaction with the robot is burdensome, it will probably be minimized, decreasing SA as autonomy increases. Future research is needed to confirm these findings, and with empirical support, task interdependence could be a path towards increasingly autonomous robot systems.

**Acknowledgements.** The research reported in this document/presentation was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory. The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

1. Endsley, M.R., Kiris, E.O.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors* 37, 381–394 (1995)
2. Schuster, D., Jentsch, F., Fincannon, T., Ososky, S.: The Impact of Type and Level of Automation on Situation Awareness and Performance in Human-Robot Interaction. In: *HCI International, Las Vegas* (in press, 2013)
3. Endsley, M.R.: Situation Awareness Global Assessment Technique (SAGAT). In: *IEEE 1988 National Aerospace and Electronics Conference*, vol. 3, pp. 789–795 (1988)
4. Rousseau, R., Tremblay, S., Breton, R.: Defining and Modeling Situation Awareness: A Critical Review. In: Banbury, S., Tremblay, S. (eds.) *A Cognitive Approach to Situation Awareness: Theory and Application*, pp. 3–21. Ashgate, Burlington (2004)
5. Salas, E., Prince, C., Baker, D.P., Shrestha, L.: Situation Awareness in Team Performance: Implications for Measurement and Training. *Human Factors* 37, 123–136 (1995)
6. Durso, F.T., Sethumadhavan, A.: Situation awareness: Understanding Dynamic Environments. *Human Factors* 50, 442–448 (2008)
7. Parasuraman, R., Cosenzo, K.A., De Visser, A.E.: Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology* 21, 270–297 (2009)
8. Johnson, M., Bradshaw, J.M., Feltovich, P.J., Hoffman, R.R., Jonker, C., Van Riemsdijk, B., Sierhuis, M.: Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design. *IEEE Intelligent Systems* 26(3), 81–88 (2011)
9. Georgeopolousis, B.S.: *Organizational Structure, Problem Solving, and Effectiveness*. Jossey-Bass, San Francisco (1986)
10. Cummings, T.G.: Self-Regulating Work Groups: A Socio-Technical Synthesis. *The Academy of Management Review* 3(3), 625–634 (1978)
11. Thompson, J.D.: *Organizations in Action*. McGraw-Hill, New York (1967)
12. Staudenmayer, N.: *Interdependency: Conceptual, Empirical, & Practical Issues*. Technical report. Massachusetts Institute of Technology (1997)
13. Wageman, R.: The meaning of interdependence. In: Turner, M.E. (ed.) *Groups at Work*, pp. 197–217. Lawrence Erlbaum Associates, Mahwah (2001)
14. Langfred, C.W.: Autonomy and Performance in Teams: The Multilevel Moderating Effect of Task Interdependence. *Journal of Management* 31, 513–529 (2005)
15. Van Der Veegt, G., Van De Vliert, E.: Intragroup Interdependence and Effectiveness: Review and Proposed Directions for Theory and Practice. *Journal of Managerial Psychology* 17, 50–67 (2002)
16. Langfred, C.W., Shanley, M.T.: Small Group Research: Autonomous Teams and Progress in Issues of Context and Levels of Analysis. In: Golembiewski, R. (ed.) *Handbook of Organizational Behavior*, 2nd edn., pp. 81–111. Marcel Dekker, New York (2001)
17. Saavedra, R., Earley, P.C., Van Dyne, L.: Complex Interdependence in Task-Performing Groups. *Journal of Applied Psychology* 78(1), 1 (1993)
18. LePine, J.A., Piccolo, R.F., Jackson, C.L., Mathieu, J.E., Saul, J.R.: A Meta-Analysis of Teamwork Processes: Tests of a Multidimensional Model and Relationships with Team Effectiveness Criteria. *Personnel Psychology* 61, 273–307 (2008)
19. Liden, R.C., Wayne, S.J., Bradway, L.K.: Task Interdependence as a Moderator of the Relationship between Group Control and Performance. *Human Relations* 50(2), 169–181 (1997)

20. Johnson, M., Bradshaw, J.M., Feltovich, P.J., Jonker, C.M., van Riemsdijk, B., Sierhuis, M.: The Fundamental Principle of Coactive Design: Interdependence Must Shape Autonomy. In: De Vos, M., Fornara, N., Pitt, J.V., Vouros, G. (eds.) COIN 2010. LNCS, vol. 6541, pp. 172–191. Springer, Heidelberg (2011)
21. Stewart, G.L., Barrick, M.R.: Team Structure and Performance: Assessing the Mediating Role of Intrateam Process and the Moderating Role of Task Type. *The Academy of Management Journal* 43(2), 135–148 (2000)
22. Riley, J.M., Strater, L.D., Chappell, S.L., Connors, E.S., Endsley, M.R.: Situation Awareness in Human-Robot Interaction: Challenges and User Interface Requirements. In: Barnes, M., Jentsch, F. (eds.) *Human-Robot Interactions in Future Military Operations*, pp. 171–191. Ashgate, Surrey (2010)
23. Wickens, C.D., Li, H., Sebok, A., Sarter, N.B.: Stages and Levels of Automation: An Integrated Meta-Analysis. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, pp. 389–393 (2010)

# Cybernetic Teams: Towards the Implementation of Team Heuristics in HRI

Travis J. Wiltshire<sup>1</sup>, Dustin C. Smith<sup>2</sup>, and Joseph R. Keebler<sup>2</sup>

<sup>1</sup> University of Central Florida, Orlando, Florida  
twiltshi@ist.ucf.edu

<sup>2</sup> Wichita State University, Wichita, Kansas  
{joseph.keebler, dcsmith}@wichita.edu

**Abstract.** This paper examines a future embedded with “cybernetic teams”: teams of physical, biological, social, cognitive, and technological components; namely, humans and robots that communicate, coordinate, and cooperate as teammates to perform work. For such teams to be realized, we submit that these robots must be physically embodied, autonomous, intelligent, and interactive. As such, we argue that use of increasingly social robots is essential for shifting the perception of robots as tools to robots as teammates and these robots are the type best suited for cybernetic teams. Building from these concepts, we attempt to articulate and adapt team heuristics from research in human teams to this context. In sum, research and technical efforts in this area are still quite novel and thus warranted to shape the teams of the future.

**Keywords:** Human-robot interaction, team heuristics, cybernetic teams, social robots.

## 1 Introduction

Human-robot interaction is a rapidly expanding multidisciplinary field that will evolve significantly over the course of the next half-century. In particular, there will not only be an increase in unmanned vehicle usage [1], both internationally and domestically, but also an increase in artificial intelligence (AI) leading to increasingly interactive and autonomous robotic systems. These advances may lead to a multitude of Human Factors issues, ranging from the “how and when” of implementing intelligent robotic systems, to the creation of dynamic human-robot (HR) teams. Previous work has demonstrated a need for adapting human team-training heuristics to HR teams [2]. Specifically, concepts such as supporting precise and accurate communication, diagnosing communication errors, providing practice opportunities, and building team orientation need to be realized in the context of HR teams. Though this is a demanding challenge, it leaves the area ripe for research. There is much to be done to understand the nature of HRI, when robots are no longer perceived as tools, but instead as team members. Therefore, this paper will focus on some of the aspects of human teams that may translate readily to HR teams. Specifically, team heuristics [3], such as those detailed above, will be examined as a strong foundational starting point.



Specifically, this paper will review relevant team heuristics theories in light of how they may be applied to future HR teams.

However, prior to reviewing these team heuristics, we first define our reconceptualization of a team by elaborating on what we mean by cybernetic teams. Then, we set the stage for how robots can even begin to possess the capabilities and intelligence required for fulfilling the role of a teammate by comparing human-agent and human-robot teams. We next elaborate on the specific types of robots that have been argued to be most capable of performing as effective teammates; namely, social robots [e.g., 4].

Although robots with such capabilities are not a widespread technology that have been instantiated as team members in operational environments, it is pertinent to uncover the issues discussed in this paper now in order to aid in the guidance of design. We may be decades away from social robots as teammates, but by applying our current knowledge of human team cognition and behavior to what we believe these cybernetic teams will be like in the future can only lead to a fortuitous and better pre-conceptualized endeavor.

## 1.1 Cybernetic Teams

With this paper, we first aim to open a discussion for understanding these “cybernetic” teams long before they are realized as technological instantiations. By cybernetic teams, we mean that the physical, biological, cognitive, and social systems traditionally comprising human teams must be extended through the further consideration of the mechanical, electronic, and technological constituents of the cybernetic team system. We argue that only from this broader perspective and through the careful consideration of the interdependent relations of each facet of these teams can the state of the art in HR teaming truly be advanced. Specifically, this allows for the necessary reconceptualization of robots as not just tools essential for completing a given task, but rather as a teammate that can interact dynamically and autonomously under varying conditions. Though certain aspects of human teams are still quite relevant to cybernetic teams, there will undoubtedly be emergent changes when HR teams are integrated more thoroughly through increased AI capabilities. In order to better explicate the nature of cybernetic teams, in the section following we draw a distinction between human-agent teams and human-robot teams.

## 1.2 Comparing Human-Agent Teams with Human-Robot Teams

Human-agent team research is one area of the literature that likely has established the groundwork for HR teams. The difference between human-agent teams and HR teams likely depends on the physical embodiment or presence of the robot in the real-world as opposed to a virtual world. Relatedly, Sukthankar, Shumaker, and Lewis [5] distinguish between a software agent and an embodied agent. *Software agents* are artificially intelligent systems that serve as intelligent members of a team, but are not necessarily tangible. They carry out algorithmic functions with regard to digital information rather than physically manipulating the world. One common example of a software agent is Apple’s Siri, which essentially is just a voice from which a tangible concept or even image of the software agent is difficult to extract. An *embodied*

*agent*, on the other hand, refers to the tangible entity that is able to carry out physical tasks and algorithmically deduce the best methods to carry out those functions. Here, we acknowledge that some embodied agents inhabit a strictly virtual environment; however, for our purposes we consider embodied agents to be those who inhabit the same physical environment as humans do.

In the context of human-robot teams, most commonly, software agents are embedded within and facilitate the operation of unmanned vehicles; however, given a certain degree of autonomy, intelligence, and dynamic interactive capabilities such unmanned vehicles could be considered robotic teammates [6]. In support of this notion, Suktankar, Shumaker, and Lewis [5] later point out that if the AI of an agent acts as an extension of the human operators (i.e. augmenting cognition), then the team does not qualify as a pure human-agent team. The key difference here between human-agent teams and human-robot teams can thus be said to reside in both the intelligence and the interactive capabilities of the agents. This, to some degree, is what we are referring to as a cybernetic team, in which, the defining criteria is the dynamic and interactive collaboration with an intelligent agent as opposed to a unidirectional utilization of the agent. That is, team decision making and other team processes become more conversational and bi-directionally interactive, rather than purely command based. Therefore, it appears as though, from a foundation of human-agent teams, emerges the key distinction of human-robot teams. That is, HR teams rely on the use of autonomous, interactive, and physically embodied intelligent robots. From here we use HR teams and cybernetic teams interchangeably. However, it would be misleading to leave our discussion of robot teammates at this point, as there is much to unpack in regards to what it means for a robot to be able to autonomously interact with human teammates. As such, our next section aims to address this very point, by detailing advances in social robotics, that aim to provide robots with precisely these capabilities.

### **1.3 Social Robots as the Enabling Factor for Human-Robot Teaming**

For robots to be successful teammates, it is essential to consider the degree to which such entities will be embodied and embedded in an information rich and complex social environment [7]. By this, we mean that it is naïve to envision robotic teammates working with humans without any sort of social intelligence or interactive capabilities. Specifically, it has been argued that effective human-robot teaming may only be achieved when robots have gained the appropriate social intelligence that allows them to function both naturally and intuitively in social interactions with humans [8-9]. That is, only when given this type of capability will robot teammates be able to work with humans towards shared goals and dynamically adjust plans based on the observation of human actions and the inferred social implications [10]. As such, further description of such robots is warranted to explicate what is needed for robots to function as teammates.

Social robots have previously been defined as robots that are: (1) physically embodied agents that, (2) function with a least some degree of autonomy, and are (3) capable of interacting and communicating with humans by, (4) adhering to normative and expected social behaviors [11]. Elaborating on this, [12] describe socially interactive robots as those that are able to (1) express or perceive emotions, (2) use high-level dialogue for communication, (3) have the ability to learn and recognize other agents,

(4) establish and maintain social relationships, and (5) use and perceive natural cues such as gaze and gestures. Depending on the domain and task, some or all of the aforementioned characteristics of social and socially interactive robots may be necessary. In instances where, for example, a robot must collaborate in a complex and high-stakes environment such as on the international space station, the bi-directional and dynamic features of collaborative work necessitates that the robot possess these social skills [10]. Accordingly, our aim here is not to review in detail the impressive advances in social robotics over the past decade; rather, our aim is to emphasize that these efforts are essential to the design of robots that will in turn facilitate effective human-robot teaming. Next, we describe the specific team heuristics from extensive research in human-human teams and describe how these will be useful in the context of HR teams.

## 2 Teaming Heuristics

Many important aspects of human teams have been established by work explicating team heuristics or, in other words, guidelines for successful team work [3]. Specifically, applications of team heuristics that ensure the team is working cohesively have been shown to substantially benefit team outcomes [13]. Therefore, it is important, we argue, to apply team heuristics to the design of robotic systems that will serve as team members in order to enable such robots to communicate, coordinate, and cooperate *as if* they were human teammates. Future robotic systems will need to “understand” these heuristics, and be able to adapt to the needs of the team based on these principles. For example, a heuristic such as ‘update the plan’ becomes complicated with the addition of a robotic team member: Which modality of communication does the robot use? How often does it need to update the plan based on its programming? Which team members need to be made aware of the updates? This leads to further design implications: Which type of communication will robotic assets be able to use? Which types of communication should they use? On human-only teams, some of these issues are solved through implicit and explicit communication, so how can we best integrate HR teams to have effective implicit and explicit communication? As can be seen, there are many questions, yet research has not been provided much in the way of answers to these questions. Accordingly, throughout this section we provide details of the most relevant team heuristics and attempt to convey how they could be realized in cybernetic teams.

### 2.1 Use Closed-Loop Communication

The most effective form of team communication is via the method of closed-loop communication. This method of communication establishes a standard of verification in which team members (i.e. sender and receiver of information) are required to acknowledge receipt of information [3], [13]. This is integral to ensuring that the communicated message has reached its intended destination and that all parties acknowledge receipt of and understanding of the communicated information.

In the case of the current state of robotics, the modality that may be best suited for closed loop communication will likely emphasize redundancy in scenarios where

noise is a primary factor of miscommunication [14]. Therefore, in fact, it may be multi-modal communication (MMC) that is best suited for facilitating closed-loop communication. MMC has recently been defined in the context of HRI as the flexible selection and exchange of information through the blending of auditory, visual, and tactile modalities in either an explicit or implicit communication [15]. Further, “explicit communication is the purposeful conveyance of information through multiple modalities...that has a defined meaning”; whereas, “implicit communication is the inadvertent conveyance of information about emotional and contextual state that will affect interpretation, thoughts, and behaviors” [15, p. 462]. The distinction between explicit and implicit communication leads to the question of which types of communication will need to be closed loop.

It is conceivable that explicit communications given their deliberate nature are most easily adopted for closed-loop communications. However, implicit communications particularly those conveyed by humans (e.g., body language) are equally relevant for certain tasks. Accordingly, technological systems for closed-loop communications in cybernetic teams are still largely undeveloped although such a system may display text and other key features of a given task through, for example, a head-mounted augmented reality display system. Nonetheless, robot teammates will require the appropriate social intelligence to understand both explicit and implicit communication whether or not they are implemented through closed-loop communication; and further, more communication options will become increasingly available to robots as the technologies advance ultimately leading to narratively structured communications analogous to human dialogue [12]. Though such advances in closed-loop communications would help to facilitate the communication operations of any cybernetic team, the consistent diagnosis of communication errors would help to ensure both natural and resilient team performance.

## **2.2 Diagnose Communication Errors**

Due to the complex nature of cybernetic teams and the ever evolving design of robotic systems by humans, “communication errors may be at multiple levels, and may include bandwidth issues, equipment failures, as well as incongruities of robotic assets” [2]. Notably, the types of communication errors and breakdowns in cybernetic teams will likely be quite different than those in human teams. As such, it is important for robotic teammates to remain as transparent as possible when it comes to issues in communication. This is essential for two purposes. On the one hand, arising issues that could negatively affect communication during the execution of a given task need to be explicitly communicated to team members. In cases such as this, the difference between signal loss due to physical obstruction is a very different issue than signal loss due to over-burdened bandwidth. Both require entirely different solutions, yet without understanding of the system’s error, human team members may easily become frustrated and distrustful of robotic teammates.

On the other hand, diagnosis of communication errors is a task that should be continually examined by the designers and engineers of these robotic systems to ensure an ongoing mitigation of these errors thus improving the overall performance of the cybernetic team. Traditional post-mission debriefs used in human teams may be a useful strategy for the improvement of communication issues. Specifically, after a

given mission human teammates could collaborate with designers and engineers to reflect on the communication errors and identify opportunities for correcting such issues. Ultimately, as machine learning and cognitive architectures for robotic systems advance, these robotic teammates will become increasingly metacognitive and self-corrective on their own; though this is certainly far from being instantiated, efforts are underway to provide robots with such capabilities [e.g., 16].

### **2.3 Evenly Distribute Workload Proportionally to Expertise**

Salas et al. [3] emphasize the importance of utilizing the skillsets of each team member regardless of their seniority. For the scope of this paper, we will consider the distribution of workload with regard to a cybernetic team, although some research in HRI has examined the results of team performance when robots are assigned a more senior role [see 17]. As autonomy increases the amount and type of work executable by robotic teammates will evolve. That is, the workload for robotic teammates will change from a monotonous and repetitive task role to an increasingly dynamic and open ended role. Thus, traditionally, robots and machines have been more suited to conduct tasks or functions such as working for long hours without rest or conducting mundane tasks; however, as the technologies advance careful attention will need to be paid in selection and designation of tasks and workload to either the human or the robotic teammate.

On the other hand, robots with the appropriate social intelligence are more likely to adaptively respond to the shifting needs of their human teammates. By this we mean that give appropriate social-cognitive mechanisms, these robotic teammates would be able to not only interpret but also predict the intentions and thus the actions of human teammates in order to interact dynamically and share the workload for a given task [9], [18]. Such mechanisms have shown to be essential for effective coordination between humans and teammates [18]. Quite to the contrast, most humans do not interact and are not familiar with robots, which leads us to our next team heuristic.

### **2.4 Frequent Practice Opportunities**

The importance of practice remains constant across human teams and cybernetic teams. In fact, it may be particularly more relevant for cybernetic teams given the novelty of interaction with robots. Specifically, it has been argued that “practice for HR teams should be frequent and mandatory. Practicing communication, missions, etc. will only enhance team performance” [2]. Practice in this sense can serve as a bi-directional benefit to human and robot teammates. On the one hand, humans gain familiarity working with the robot and perceiving it as a teammate and in doing so begin to develop trust in the system and fluidity in the types of interaction, among other things. In the case of the robot, it may need an interaction period of a certain duration in which it can learn about the behaviors of the human in order to begin to coordinate as an effective teammate. Of course this depends on the types of intelligence it is programmed with, but it is likely the benefit of practice remains constant. In short, practice provides an opportunity for missions and tasks to be rehearsed in contexts in which the stakes are not high such that, the chances of success in complex operations are improved. Practice and interaction more generally between human and robot teammates can also lead to benefits in the convergence of mental models.

## 2.5 Refine Shared Mental Models

Prior research on shared mental models, within the context of HR teams has examined the importance of the convergence of mental models for enhanced team performance [19]. If mental models converge with flawed content, team performance can be negatively affected, resulting in situation assessment errors and conflict within the team [20-22]. In light of this, it is suggested that future research examine ways to decrease flawed mental model convergence in order to enhance team performance. Specifically, in the context of HR teams, we must ensure that the human teammate's mental model is properly suited to the dynamics of the robotic teammate. In particular, efforts are also needed to further explore the ways in which the notion of mental models can be instantiated in such robot systems [see 23].

As detailed previously, HRI researchers have recognized the importance of explicit and implicit communication between humans and robots [2], [15]. That being said, integrating robots into human teams will require both types of communication, and could therefore, increased communications between humans could decrease errors in mental model convergence. Furthermore, through combination of closed-loop communication augmented with multi-modal communication systems as well as appropriate social intelligence, it is expected that adequate mental model convergence would ensue thus facilitating efficient teamwork. Of course efforts are needed to not only instantiate the notion of mental models in robots but to empirically examine the effects of such efforts and the variables that play a role in both their accurate and inaccurate convergence.

## 2.6 Manifest Deep Understanding of Tasks

Typically the emphasis here is on a deliberate intervention in which a designated team leader encourages team members to provide environmental assessments to better define the tasks and situation parameters leading to the creation of well-developed plans and the development of adaptive expertise in team members [3]. Thorough explication of these issues prior to practice sessions or missions can help to enable the coordination and success of the team. This provides each team member with more detailed and flexible understandings of the dynamic nature of their tasks. However, cybernetic teams will not only require such interventions for effective team performance, but also for understanding the capabilities of robotic teammates. Software updates will likely be relentless, of course, in pursuit of better robotic teammates, but nonetheless, often a game changer. Once robots reach a certain level of intelligence and interactivity the modifications of their software will be limitless and could also at some point become self-corrective. Thus, an adaptive expertise in terms of interaction with robotic teammates must be developed for the assimilation of new software and how that affects future team operations. This notion is related to our next team heuristic.

## 2.7 Build Team Orientation

Given the novelty of robotic teammates it is essential to “integrate robots early in team formation so that roles can be discovered and trust established” [2]. Trust in robotic teammates will not happen overnight; however, it essential for cooperation.

More so, if humans are untrusting or frustrated with the performance of their robotic teammates they will be less likely to cooperate and this could result in putting the team at risk of failure. Relatedly, depending on the intelligence and programming of the robot, instances could result in humans and robotic teammates that hold divergent viewpoints. Thus, through team orientation these cybernetic teams can become familiar with the varying perspectives and functions of each team member in relation to a given task. Early and frequent team orientation is thus recommended and can be instantiated through required practice and informal interactions. Benefits are likely to include improvements to team performance in the field by giving the team a chance to interact in a non-stressful environment when stakes are low. This will give the human team members time to understand the capabilities of their robotic teammate(s), as well as allow for the robotic asset to socially engage the team and build rapport.

### 3 Conclusions

In sum, we have first examined a reconceptualization of the traditional notion of a team of which we have termed *cybernetic teams*. We argue that this reconceptualization allows for greater consideration and treatment of the physical, biological, social, cognitive, mechanical, electronic, and technological components of such teams as a unified system. As these types of teams become increasingly prevalent, such a reconceptualization is necessary to foster better designs and ways of improving team processes and performance without neglecting any element of such a complex interdependent system. Next, we have drawn from human-agent teams to attempt a clear articulation at what is meant by a human-robot team. That is, a team in which robots are physically embodied, autonomous, interactive, and intelligent. However, one of the key contributions here is that for robots to ever be thought of as teammates, they must possess the appropriate social intelligence and interactive capabilities that allow them to function intuitively and naturally with human teammates. Building on this, we have reviewed and adapted some of the team heuristics that stem from the study of human teams and attempted to articulate how these might be realized in cybernetic teams. Of course, it is likely that as the state of the art advances in such teams, novel cybernetic team heuristics will emerge. Nonetheless, efforts such as this as well as empirical and technical efforts are warranted to instantiate and evaluate robots with the capabilities discussed herein and as a result, develop the teams of the future.

**Acknowledgements.** This work was partially supported by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, the U.S. Government or the University of Central Florida. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## References

1. Clapper, J.R., Young, J.J., Cartwright, J.E., Grimes, J.G.: FY 2009-2034 unmanned systems integrated roadmap. Department of Defense: Office of the Secretary of Defense Unmanned Systems Roadmap (2009)
2. Keebler, J.R., Jentsch, F., Fincannon, T., Hudson, I.: Applying team heuristics to future human-robot systems. In: Proceedings of Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 169–170 (2012)
3. Salas, E., Wilson, K.A., Murphy, C.E., King, H., Salisbury, M.: Communicating, coordinating, and cooperating when lives depend on it: Tips for teamwork. *The Joint Commission Journal on Quality and Patient Safety* 34, 333–341 (2008)
4. Breazeal, C.: Role of expressive behavior for robots that learn from people. *Philosophical Transactions of the Royal Society: Biological Sciences* 365(1353), 3527–3538 (2009)
5. Sukthankar, G., Shumaker, R., Lewis, M.: Intelligent agents as teammates. In: Salas, E., Fiore, S.M., Letsky, M.P. (eds.) *Theories of Team Cognition: Cross-disciplinary Perspectives*, pp. 313–343. Taylor & Francis Group, New York (2012)
6. Syraca, K., Lewis, M.: Integrating agents into human teams. In: Salas, E., Fiore, S.M. (eds.) *Team Cognition: Understanding the Factors that Drive Process and Performance*. American Psychological Association, Washington, DC (2004)
7. Dautenhahn, K., Ogden, B., Quick, T.: From embodied to socially embedded agents – Implications for interaction-aware robots. *Cognitive Systems Research* 3(3), 397–428 (2002)
8. Breazeal, C.: Social interactions in HRI: The robot view. *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews* (2004)
9. Streater, J.P., Bockelman Morrow, P., Fiore, S.M.: Making things that understand people: The beginnings of an interdisciplinary approach for engineering computational social intelligence. Presented at the 56th Annual Meeting of the Human Factors and Ergonomics Society, Boston, MA, October 22-26 (2012)
10. Hoffman, G., Breazeal, C.: Collaboration in human-robot teams. In: Proceedings of AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, pp. 1–18. AIAA, Reston (2004)
11. Bartneck, C., Forlizzi, J.: A designed-centered framework for social human-robot interaction. In: Proceedings of the Ro-Man 2004, Kurashiki, pp. 591–594 (2004)
12. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 143–166 (2003)
13. Weaver, S.J., Rosen, M.A., Diaz-Grandados, D., Lazzara, E.H., Lyons, R., Salas, E., Knych, S.A., McKeever, M., Adler, L., Barker, M., King, H.B.: Does teamwork improve performance in the operating room? A multilevel evaluation. *The Joint Commission Journal on Quality and Patient Safety* 36, 133–142 (2010)
14. Weaver, S.J., Feitosa, J., Salas, E., Seddon, R., Vozenilek, J.A.: The theoretical drivers and models of team performance and effectiveness for patient safety. In: Salas, E., Frush, K. (eds.) *Improving Patient Safety Through Teamwork and Team Training*, Oxford Press, New York (2013)
15. Lackey, S., Barber, D., Reinerman, L., Badler, N.I., Hudson, I.: Defining next-generation multi-modal communication in human robot interaction. In: Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting, pp. 460–464 (2011)
16. Kurup, U., Lebiere, C.: What can cognitive architectures do for robotics? *Biologically Inspired Cognitive Architectures* 2, 88–99 (2012)
17. Hinds, P.J., Roberts, T.L., Jones, H.: Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human Computer Interaction* 19(1&2), 151–181 (2004)



18. Breazeal, C., Gray, J., Berlin, M.: An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research* 28, 656–680 (2009)
19. Stiefelhagen, R., Ekenel, H.K., Fugen, C., Gieselmann, P., Holzapfel, H., Kraft, F., Nickel, K., Voit, M., Waibel, A.: Enabling multimodal human–robot interaction for the Karlsruhe humanoid robot. *IEEE Transactions on Robotics* 23(5), 840–851 (2007)
20. McComb, S.: Shared mental models and their convergence. In: Letsky, M.P., Warner, N.W., Fiore, S.M., Smith, C.A.P. (eds.) *Macro-cognition in Teams: Theories and Methodologies*, pp. 35–50. Ashgate Publishing, Aldershot (2008)
21. McComb, S., Vozdolska, R.: Capturing the convergence of multiple mental models and their impact on team performance. Meeting of the Southwest Academy of Management, San Diego, CA (2007)
22. McComb, S., Kennedy, D.: Facilitating effective mental model convergence: the interplay among the team’s task, mental model content, communication flow, and media. In: Salas, E., Fiore, S.M., Letsky, M.P. (eds.) *Theories of Team Cognition: Cross-disciplinary Perspectives*, pp. 549–570. Taylor & Francis Group, New York (2012)
23. Schuster, D., Ososky, S., Jentsch, F., Phillips, E., Lebiere, C., Evans, A.W.: A research approach to shared mental models and situation assessment in future robot teams. In: *Proceedings of the 55th Annual Meeting on Human Factors and Ergonomics Society*, pp. 456–460 (2011)

## **Part IV**

# **Presence and Tele-Presence**

# Embodiment and Embodied Cognition

Mark R. Costa<sup>1</sup>, Sung Yeun Kim<sup>2</sup>, and Frank Biocca<sup>2</sup>

<sup>1</sup> School of Information Studies,  
<sup>2</sup> S.I. Newhouse School of Public Communications,  
Syracuse University M.I.N.D. Lab.  
{mrcosta, skim154, fbiocca}@syr.edu

**Abstract.** Progressive embodiment and the subsequent enhancement of presence have been important goals of VR researchers and designers for some time (Biocca, 1997). Consequently, researchers frequently explore the relationship between increasing embodiment and presence yet rarely emphasize the ties between their work and other work on embodiment. More specifically, we argue that experiments manipulating or implementing visual scale, avatar customization, sensory enrichment, and haptic feedback, to name a few examples, all have embodiment as their independent variable. However, very few studies explicitly frame their work as an exploration of embodiment. In this paper we will leverage the field of Embodied Cognition to help clarify the concept of embodiment.

**Keywords:** human-computer interaction, presence, embodied cognition, virtual reality.

## 1 Introduction

Increasing presence is one the primary goals of virtual reality (VR) researchers and developers, whether it is intended to improve the entertainment value of the experience or ability to affect change in the user's real world behavior. A significant portion of the research and development into improving presence is centered on sensory stimulation, avatar mobility, and avatar representativeness. In the past researchers argued that these affordances all fall under embodiment (Biocca, 1997), but are in practice are rarely treated as related ideas. Despite there being a seemingly visible relationship between these research areas as well as a tentative label, very little effort has been put into clearly defining the concept linking these loosely tied areas together.

In order to resolve this problem we put forth the concept of afforded embodiment, or the degree to which the avatar provides equal or greater functionality than the user's natural body. We draw on recent developments in the area of embodied cognition to support an argument that VR researchers need to pay much closer attention to the relationship between avatar functionality and the activities within the virtual environment. The foundation of this argument is based on the basic premise of embodied cognition, which is that the body plays a constituent, not causal role in cognition (Shapiro, 2011). From this perspective, avatars with limited functionality limit their user's ability to mentally explore the virtual environment. We argue that a clearer

understanding of the role of the body in cognition will allow researchers and designers to better prioritize design considerations for avatar functionality based on the activities or goals of the VR experience.

## 2 Background and Literature Review

### 2.1 Embodied Cognition

Embodied cognition (EC) is a research program that re-focuses cognitive science research to include the body as a critical component of cognition (Shapiro, 2007). EC assumes that our reality is shaped by the interactions of our mind, body, and environment. This contrasts with those who would argue that reality is shaped by our mind and the mind manipulates the body through abstract symbols (Shapiro, 2007). While EC is not a predictive theory, it gives virtual reality (VR) researchers an opportunity to revisit and clarify a neglected concept within our field as well as its relationship to presence.

There are two types of experimental evidence for embodied cognition - behavioral and neurological. We present the neurological evidence first and interpret the findings based on the assumption that objects and movements that activate the same neuronal systems are “linked” by the brain. For example, when examining neurons responsible for manipulating objects, objects of similar size and shape will activate the same set of neurons for a specific set of hand motions even when the subject is not actually manipulating the object, while objects of different shapes and sizes will activate a different set of neurons, even when the actual movements to manipulate the object are the same. In a subsequent study, a set of inferior premotor neurons were found to be responsible for executing a limited set of distal arm movements. However, that same set of neurons will activate if an observance within the environment is related to the physical process overseen by the neural network, even if there is no motor activity present (Gazzola & Keysers, 2009; Pellegrino et al., 1992). The cognitive representation of the object is intrinsically tied to the way the subject uses its body to manipulate the object. They are not two separate symbols pieced together by the mind for action, but are bound together in the same schemata (Rizzolatti et al., 1988).

The behavioral examples are more straightforward and demonstrate a clear link between either sensory perception and cognitive perception or motor action and cognitive perception. For example, researchers found that by priming individuals with a warm or cold beverage they could influence whether or not the subject perceived those they just met as having a warm or cold personality, respectively (Williams & Bargh, 2008). In an unrelated study, researchers found that sitting upright influences the extent to which subjects felt pride in an achievement. A second study found that contraction of the forehead muscles influenced subjects’ perceptions of how hard they worked on a task (Stepper & Strack, 1993).

There is also a demonstrated link between body movements and improved problem solving performance. Subjects who used appropriate body motions during a physics problem solving session performed significantly better than subjects who used inappropriate body motions (Thomas & Lleras, 2009). In another study, researchers found that preventing the facial expression aligned to the sentiment of a sentence significantly impacted the subjects’ reading performance (Havas et al., 2010).

The previous examples reveal the extent to which the body, mind, cognitive, and emotional states are all intertwined. We create mental schemata of environmental objects via the way in which we manipulate them (Pellegrino et al., 1992; Rizzolatti et al., 1988) and encode our own bodily movements and perceive others' movements using the same set of neurons (Gazzola & Keysers, 2009). Behaviorally, our body positioning and use impacts our social observations (Williams & Bargh, 2008), feedback acceptance and task performance recall (Stepper & Strack, 1993), and improves performance on cognitive tasks (Havas et al., 2010; Thomas & Lleras, 2009). Consequently, it is critically important for VR designers and researchers to acknowledge the many subtle, yet fundamental, ways in which our body influences our perceptions, actions, and emotions.

## 2.2 Embodiment

Embodiment is a concept used in many different research areas and thus has many different, yet related meanings. In this section we will briefly discuss the different meanings and uses of embodiment, identify which of the variations we chose to adopt for this research, how we operationalize the concept, and our justifications for doing so.

Damasio (1994) and other embodied cognition theorists argue that the body serves as the central framework for our interactions with the world. We perceive the physical world in relation to our body, and therefore what we know about the world is “constructed from patterns of energy detected by the body” (Biocca, 1997). The body is also a communication device and is a critical tool for expressing mental states (Benthall & Polhemus, 1975). The implications of this theory for VRE researchers and designers is that in order for the user to incorporate the VE into his/her reality, the system must provide affordances for users so that they may become embodied, or take some form or shape within the VRE.

Ziemcke (2003) identified six different uses of the term embodiment across multiple streams of research. Of the six notions of embodiment addressed by Ziemcke, structural coupling, historical coupling, and social embodiment are the most relevant.

Structural coupling is the notion that organisms are embodied in their environment if actions by one affect the other. Quick and colleagues (Quick, 1999 as cited in Ziemcke, 2003) articulate this idea clearly, saying that “*A system X is embodied in an environment E if perturbatory channels exist between the two.*” More concretely, we can say a user is embodied in a virtual reality environment (VRE) if changes in the VRE affect the user, and the user can affect the VRE. Historical structural coupling is an extension of the idea of structural coupling in that it argues embodiment increases through a series of interactions between the system and the environment.

Ziemcke (2003) viewed social presence as orthogonal to the other definitions of presence because it addresses the role of embodiment in social situations, versus what kind of body is required for different situations. Barsalou et al. (2003) describe social embodiment as “states of the body, such as postures, arm movements, and facial expressions, [which] arise during social interaction and play central roles in social information processing (Ziemcke, 2003).” Mennecke and colleagues (2010) formally incorporate this idea into their Embodied Social Presence Theory, arguing that the body is the nexus of communication and that embodied representations combined

with goal-directed shared activities affects the perceptions of users by drawing them into higher levels of cognitive engagement.

In contrast to Ziemke's 6 different uses of embodiment, Lakoff and Johnson (1999) identify 3 distinct levels of embodiment - neural, cognitive unconscious, and phenomenological conscious experience.

Biocca's (1997) work on embodiment in VRE draws on Damasio's (1994) ideas of how the human brain constructs reality through interactions of the body with the environment, and fits within the notions of structural coupling identified by Ziemke (2003).

While Biocca provides no formal definition of embodiment, it is clear that he views progressive embodiment as providing increasingly "natural" functionality to avatars. Natural functionality includes sensory perception in the form of high fidelity audio or visual stimulation, haptic feedback, natural motion control etc. It is also important to note that our bodies fill an important role in social interaction and self-identification, and therefore must provide increasing affordances for how we expect to use our bodies in those situations as well.

Embodiment is expected to have a direct impact on various forms of presence (Biocca, 1997). With increasing embodiment we expect increasing levels of psychophysiological responses to VE. The brain's relationship to the body is highly malleable; therefore it is possible to convince the brain that it will suffer the consequences of actions within the VE. Biocca (1997) addresses this idea in his three-way relationships between our brain's mental models of our physical, virtual, and phenomenal selves. A second, longer-term implication of increasing embodiment - it may have a permanent effect on our body schema. We may have difficulty controlling what crosses over from virtual reality to natural reality.

Once again, it is important to note that the avatar plays a critical role in the social aspects of VE. For example, Yee and Bailenson (2007) found that users who were given taller avatars were more likely to negotiate from a position of power in online trading tasks, while those given shorter avatars were more likely to accept asymmetrical trades. Taylor (2010) provides a more descriptive account of the various ways in which users construct their identity through avatar customization, as well as the degree to which most users expect exclusive use of that identity. He also describes the process of how the user identifies with the avatar, and how that identification shapes his/her perception of self. This is a known phenomenon and has been exploited to encourage changes in real-world behavior (Dean et al., 2009).

### **2.3 Neural Embodiment**

The purpose of increasing embodiment is to improve the user's sense of presence. Presence is primarily measured through a subjective post hoc questionnaire, although more recent research includes physiological measurements (Guger et al., 2004; Wiederhold et al., 2002). Further advances in psycho-physiological sensors allow us to start looking at the physiological and neurological correlates of embodiment and presence.

Within cognitive neuroscience, embodiment defined as feeling situated in one's own body (Arzy et al., 2006). This is usually researched by exploring the opposite condition, which is the out-of-body experience, or disembodiment. The most direct

translation to VR research from cognitive neuroscience would be that a user feels as if they inhabited the avatar, with concomitant physiological responses to the environment and little notice of their own “real” body.

The results from several experiments highlight the potential for better understanding the neuro-cognitive basis for the relationship between cognitive embodiment and presence. Arzy and colleagues (2006) identified two separate regions of the brain that activated dependent upon whether subjects viewed a picture of a body from an embodied perspective, or a disembodied perspective. The former condition resulted in activation of the extrastriate body area (EBA), while the latter resulted in the temporo-parietal junction (TPJ). Injury to the TPJ has been linked to out-of-body experiences in other research (Blanke et al, 2004), and has been associated with certain aspects of self-processing, self-other distinction, and mental own-body imagery (Arzy et al., 2006; Ruby & Decety, 2001; Vogeley & Fink, 2003). The EBA responds to both images of bodies and body parts, imagined movement of one’s own body, and executed movements (Arzy et al., 2006; Astafiev et al., 2004; Downing et al., 2001).

A similar result was found when researchers asked subjects to imagine themselves at some location outside of their body and then perform spatial transformations on the body (Blanke and Arzy, 2005). Subjects were then asked to perform the same task with non-body images. Subsequent artificially induced interference through transcranial magnetic stimulation interfered with the former task, but not the latter. The results suggest that the TPJ is responsible for mediating spatial unity of self and body, not and that external representations of self are not treated as normal objects for spatial transformation.

Research in cognitive neuroscience on embodiment is relevant to our explorations of embodiment and presence in virtual worlds. Understanding which regions of the brain are responsible for creating the feeling of inhabiting an avatar will allow us to better measure the user’s reaction to the avatar.

### **3 Afforded Embodiment and Virtual Reality**

Taking these ideas and aspects of embodiment, we define embodiment as the degree to which an avatar affords the user equal or greater functionality expected of our natural bodies. This functionality is comprised of three dimensions – physical [motor control and environmental manipulation], sensory input, social and self-identity. We argue that researchers have been experimenting with different degrees of afforded embodiment for years yet have not really considered their research to be part of embodiment research.

Our dimensions of embodiment map well to Lee’s (K. M. Lee, 2004) three forms of presence, although the strengths of those relationships has yet to be determined. Evidence suggests that a lack of self-presence may inhibit achieving full presence (Slater, Usoh, & Steed, 1995), which would also indicate that psycho-social embodiment is always an important consideration.

In addition to progressive sensory and motor embodiment, there needs to be a high degree of sensorimotor coupling. This is defined as the “degree to which changes in body position correlate immediately and naturally with appropriate changes in sensory feedback (Biocca, 1997).” For example, a lack of coordination between the visual, vestibular, and motor systems usually results in simulator sickness.

### 3.1 Sensory Input

The sensory input dimension measures the degree to which the user can leverage the different senses during interactions with the VE. These include, but are not limited to - vision, audition, olfaction, tactician, thermoception, proprioception, etc.

The body of research on visual sensory stimulation in VE is very large; this is not surprising given the amount of emphasis our culture places on the visual. This research includes experimentation on visual scale (i.e., screen size) (Tan et al., 2004) and dimensionality (2D vs. 3D) (Bae et al., 2012) and their effects on a range of task performances and presence.

Higher quality ambient and action driven sounds (i.e., sounds appearing to originate from the source of movement) are increasingly common. A number of researchers have been working on the role of audio in virtual worlds and have found varying degrees of effect on subjective reports of subjects' level of immersion (Grimshaw, Lindley, & Nacke, 2008).

Haptic feedback is an increasingly popular affordance in virtual environments. Force feedback has been a popular option in many simulators, and is now relatively common on many joysticks. More sophisticated implementations are also popular, as researchers seek to increase users' sense of presence. Sallnäs and colleagues (Sallnäs, Rasmus-Gröhn, & Sjöström, 2000) found that implementing a force feedback mechanism had positive effects on [physical] presence and task performance. There have also been a number of successful subsequent studies looking to use haptic feedback for improving or understanding social (Bailenson & Yee, 2007; Chan, MacLean, & McGrenere, 2008) and physical (S. Lee & Kim, 2008) presence in virtual environments.

### 3.2 Motor Control

Motor control refers to the ways in which the user can control the avatar as well as the degree to which the avatar is controllable. Avatar control generally refers to the input device used to control the movements of the avatar (gamepad, keyboard, etc.). Avatars can move through VE via different paths; in early video games movement was along one or two dimensions, while newer VEs allow much greater freedom of movement. The ability to, and activity of, engaging in body movement including bending, crouching, and head pitch and yaw, affects presence (Slater et al, 1998). Researchers have also found that mapping VE locomotion to similar real body movements results in higher presence, albeit mediated by the amount of subjective association the user has with the avatar (Slater et al., 1995). Finally, users can manipulate the VE to varying extents; how the user manipulates the environment as well as the degree to which she can has an effect on presence.

### 3.3 Psycho-social Afforded Embodiment

Psycho-social afforded embodiment refers to the degree to which the user can modify and/or manipulate their avatar to reflect or express their identity. That identity can either be an idealized or accurate reflection of their identity. The avatar must facilitate, or at least not impede, the process of identity construction. Based on previous research we know that avatar customization is an important affordance and that users



expend considerable effort to customize their avatar (Ducheneaut et al, 2009; Taylor, 2002), respond and behave according to the embodiment of their avatar (Yee & Bailenson, 2007), and feel higher levels of presence when the avatars resemble themselves (Bailey, Wise, & Bolls, 2009). Psycho-social embodiment directly affects self-presence and social presence, and may serve as a moderating variable for physical presence.

## 4 Conclusions

In this paper we discussed the ways in which the interactions between our minds, bodies, and environment form the basis for our cognition. For example, we create mental schemata of environmental objects via the way in which we manipulate them (Pellegrino et al., 1992; Rizzolatti et al., 1988). We also encode our own bodily movements and perceive others' movements using the same set of neurons (Gazzola & Keysers, 2009). Behaviorally, our body positioning and use impacts our social observations (Williams & Bargh, 2008), feedback acceptance and task performance recall (Stepper & Strack, 1993), and improves performance on cognitive tasks (Havas et al., 2010; Thomas & Lleras, 2009).

The three-way relationship between mind, body and environment is the focus of the research area known as Embodied Cognition, which in turn can be leverage to guide VR researchers and designers who are interested in making VR systems that facilitate greater levels of presence. More importantly, by acknowledging that our bodies play an integral (instead of subordinate) role in our cognitive processes and that this role has multiple dimensions we can begin to explore the more subtle yet important relationships between embodiment and presence.

Looking to the future we see opportunities to not only revisit old data, but also start exploring the ways in which avatar design affordances affects users' sense of presence in virtual environments. Additionally, we now have a framework that can be used to guide psycho-physiological instrument based research on presence. For example, it may be possible to use functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG)(Hirshfield et al., 2009) to measure users' engagement, mental workload, and response inhibitions in virtual environments to see if the affordances are working as intended.

Based on the ideas of embodied cognition, we argued that afforded embodiment is an appropriate framework for exploring avatar functionality and presence. In this paper we highlighted the already large body of literature built up around exploring degrees of embodiment. While that research is a good first step, much of it did not seek to explicitly measure the relationships between the dimensions of afforded embodiment and forms of presence. Insight into these relationships will help designers and researchers make more informed design decisions when choosing avatar affordances for virtual environments. Finally, establishing a more systematic and coherent view of the relationship between the user and their avatar is a necessary first step in understanding the user's experience in virtual environments.

**Acknowledgement.** This work was supported in part by grant No. R31-10062 from the World Class University (WCU) project of the Korean Ministry of Education, Science & Technology (MEST) and the Korea National Research Foundation (NRF) through Sungkyunkwan University (SKKU Univ.). The authors sincerely thank MEST and NRF for providing valuable resources to this research project. The project was also supported in part by the Newhouse endowment awarded to Frank Biocca.

## References

1. Arzy, S., Thut, G., Mohr, C., Michel, C.M., Blanke, O.: Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *The Journal of Neuroscience the Official Journal of the Society for Neuroscience* 26(31), 8074–8081 (2006), doi:10.1523/JNEUROSCI.0745-06.2006
2. Astafiev, S.V., Stanley, C.M., Shulman, G.L., Corbetta, M.: Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature neuroscience* 7(5), 542–548 (2004), doi:10.1038/nn1241
3. Bae, S., Lee, H., Park, H., Cho, H., Park, J., Kim, J.: The effects of egocentric and allocentric representations on presence and perceived realism: Tested in stereoscopic 3D games. *Interacting with Computers* 24(4), 251–264 (2012), doi:10.1016/j.intcom.2012.04.009
4. Bailenson, J.N., Yee, N.: Virtual interpersonal touch: Haptic interaction and copresence in collaborative virtual environments. *Multimedia Tools and Applications* 37(1), 5–14 (2007), doi:10.1007/s11042-007-0171-2
5. Bailey, R., Wise, K., Bolls, P.: How avatar customizability affects children’s arousal and subjective presence during junk food-sponsored online video games. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 12(3), 277–283 (2009), doi:10.1089/cpb.2008.0292
6. Biocca, F.: The cyborg’s dilemma: embodiment in virtual environments. In: *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*, pp. 12–26. IEEE Comput. Soc (1997), doi:10.1109/CT.1997.617676
7. Blanke, O., Landis, T., Spinelli, L., Seeck, M.: Out-of-body experience and autoscopia of neurological origin. *Brain: A Journal of Neurology* 127(Pt. 2), 243–258 (2004), doi:10.1093/brain/awh040
8. Chan, A., MacLean, K., McGrenere, J.: Designing haptic icons to support collaborative turn-taking. *International Journal of Human-Computer Studies* 66(5), 333–355 (2008), doi:10.1016/j.ijhcs.2007.11.002
9. Damasio, A.R.: *Descartes’ Error: Emotion, reason, and the brain*, p. 312. Putnam, New York (1994)
10. Dean, E., Cook, S., Keating, M., Murphy, J.: Does this Avatar Make Me Look Fat? Obesity and Interviewing in Second Life. *Journal of Virtual Worlds Research* 2(2) (2009), doi:10.4101/jvwr.v2i2.621
11. Decety, J., Grezes, J.: Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences* 3(5), 172–178 (1999), doi:10.1016/S1364-6613(99)01312-1  
Ducheneaut, N., Wen, M.-H., Yee, N., Wadley, G.: Body and mind. In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 2009*, p. 1151. ACM Press, New York (2009), <http://dl.acm.org.libezproxy2.syr.edu/citation.cfm?id=1518701.1518877>

12. Gazzola, V., Keysers, C.: The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral cortex* 19(6), 1239–1255 (1991), doi:10.1093/cercor/bhn181
13. Gibbs, R.W.: Metaphor Interpretation as Embodied Simulation. *Mind & Language* 21(3), 434–458 (2006), doi:10.1111/j.1468-0017.2006.00285.x
14. Grimshaw, M., Lindley, C., Nacke, L.: Sound and Immersion in the First-Person Shooter: Mixed Measurement of the Player's Sonic Experience. *Scientific Commons*, 7 (2008), <http://en.scientificcommons.org/49379075> (retrieved)
15. Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., et al.: Heart-Rate Variability and Event-Related ECG in Virtual Environments. In: 7th Annual International Workshop on Presence, Valencia, Spain, pp. 240–245 (2004), <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.127.361> (retrieved)
16. Havas, D.A., Glenberg, A.M., Gutowski, K.A., Lucarelli, M.J., Davidson, R.J.: Cosmetic use of botulinum toxin-a affects processing of emotional language. *Psychological Science* 21(7), 895–900 (2010), doi:10.1177/0956797610374742
17. Hirshfield, L.M., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E.T., Jacob, R.J.K., Sassaroli, A., Fantini, S.: Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload. In: Schmorow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) *FAC 2009. LNCS*, vol. 5638, pp. 239–247. Springer, Heidelberg (2009), doi:10.1007/978-3-642-02812-0
18. Lakoff, G., Johnson, M.: *Philosophy in the flesh: The embodied mind and its challenge to Western thought*, p. 640. Basic Books, New York (1999), [http://books.google.com/books?hl=en&lr=&id=KbqxnX3\\_uc0C&pgis=1](http://books.google.com/books?hl=en&lr=&id=KbqxnX3_uc0C&pgis=1)
19. Lee, K.M.: Presence, Explicated. *Communication Theory* 14(1), 27–50 (2004), doi:10.1111/j.1468-2885.2004.tb00302.x
20. Lee, S., Kim, G.J.: Effects of haptic feedback, stereoscopy, and image resolution on performance and presence in remote navigation. *International Journal of Human-Computer Studies* 66(10), 701–717 (2008), doi:10.1016/j.ijhcs.2008.05.001
21. Mennecke, B.E., Triplett, J.L., Hassall, L.M., Conde, Z.J.: Embodied Social Presence Theory. In: 2010 43rd Hawaii International Conference on System Sciences, pp. 1–10. IEEE (2010), doi:10.1109/HICSS.2010.179
22. Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G.: Understanding motor events: a neurophysiological study. *Experimental Brain Research* 91(1) (1992), doi:10.1007/BF00230027
23. Quick, T., Dautenhahn, K.: On Bots and Bacteria: Ontology Independent Embodiment. In: Floreano, D., Mondada, F. (eds.) *ECAL 1999. LNCS*, vol. 1674, pp. 339–353. Springer, Heidelberg (1999), <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.9469>
24. Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., Matelli, M.: Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research* 71(3), 491–507 (1988), doi:10.1007/BF00248742
25. Ruby, P., Decety, J.: Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature neuroscience* 4(5), 546–550 (2001), doi:10.1038/87510
26. Sallnäs, E.-L., Rasmus-Gröhn, K., Sjöström, C.: Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction* 7(4), 461–476 (2000), doi:10.1145/365058.365086

27. Shapiro, L.A.: Embodied Cognition (Google eBook), p. 237. Taylor & Francis (2011), <http://books.google.com/books?hl=en&lr=&id=Msi50Zek9vwC&pgis=1> (retrieved)
28. Slater, M., Steed, A., McCarthy, J., Maringelli, F.: The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40(3), 469–477 (1998), doi:10.1518/001872098779591368
29. Slater, M., Usoh, M., Steed, A.: Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction* 2(3), 201–219 (1995), doi:10.1145/210079.210084
30. Smith, L.B.: Cognition as a dynamic system: Principles from embodiment. *Developmental Review* 25(3–4), 278–298 (2005), doi:10.1016/j.dr.2005.11.001
31. Stepper, S., Strack, F.: Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology* 64(2), 211–220 (1993)
32. Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R.: Physically large displays improve path integration in 3D virtual navigation tasks. In: *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI 2004*, pp. 439–446. ACM Press, New York (2004), doi:10.1145/985692.985748
33. Taylor, T.L.: Living digitally: Embodiment in virtual worlds. In: Schroeder, R. (ed.) *The social life of avatars: presence and interaction in shared virtual environments*, pp. 40–62. Springer, London (2002)
34. Thomas, L.E., Lleras, A.: Swinging into thought: directed movement guides insight in problem solving. *Psychonomic bulletin & review* 16(4), 719–723 (2009), doi:10.3758/PBR.16.4.719
35. Vogeley, K., Fink, G.R.: Neural correlates of the first-person-perspective. *Trends in cognitive sciences* 7(1), 38–42 (2003), <http://www.ncbi.nlm.nih.gov/pubmed/12517357> (retrieved)
36. Wiederhold, B.K., Jang, D.P., Kim, S.I., Wiederhold, M.D.: Physiological Monitoring as an Objective Tool in Virtual Reality Therapy. *CyberPsychology & Behavior* 5(1), 77–82 (2002), doi:10.1089/109493102753685908
37. Williams, L.E., Bargh, J.A.: Experiencing physical warmth promotes interpersonal warmth. *Science* 322(5901), 606–607 (2008), doi:10.1126/science.1162548
38. Yee, N., Bailenson, J.: The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research* 33(3), 271–290 (2007), <http://doi.wiley.com/10.1111/j.1468-2958.2007.00299.x> (retrieved)
39. Ziemke, T.: What's that thing called embodiment. In: *Proceedings of the 25th Annual meeting of the Cognitive Science Society*, pp. 1305–1310 (2003)

# DigiLog Space Generator for Tele-Collaboration in an Augmented Reality Environment

Kyungwon Gil, Taejin Ha, and Woontack Woo

KAIST UVR Lab., Daejeon 305-701, S. Korea  
{kgil, taejinha, wwoo}@kaist.ac.kr

**Abstract.** Tele-collaboration can allow users to connect with a partner or their family in a remote place. Generally, tele-collaborations are performed in front of a camera and screen. Due to their fixed positions, these systems have limitations for users who are moving. This paper proposes an augmented-reality-based DigiLog Space Generator. We can generate interested space and combine remote space in real time ensuring movement. And our system uses reference object to calculate scale of space and coordinates. Scale and coordinates are saved at Database(DB) and used for realistic combination of space. DigiLog Space Generator is applicable to many AR applications. We discuss the experiences and limitations of our system. Future research is also described.

**Keywords:** Augmented Reality, Tele-collaboration, Human-Computer Interaction.

## 1 Introduction

Tele-collaboration technology aims at connecting users in different locations through a network to achieve common goals efficiently by working together. Typical tele-collaboration at present involves a user who communicates with remote participants and shares digital information in front of the screen. There are many applications for tele-collaboration: Users "attend" Web conferences with a company partner far away, using video conference or data conference systems; teachers and students look at the same data and study together in their own home or office; and so on. In recent years, tele-medicine enables a distant doctor to provide medical treatment to a patient in his or her own home. These tele-collaboration technologies can overcome the spatial limitations of traditional collaboration, because it can allow collaborators to transcend space. Moreover, tele-collaboration requires no expenditures of time and money for attending meetings.

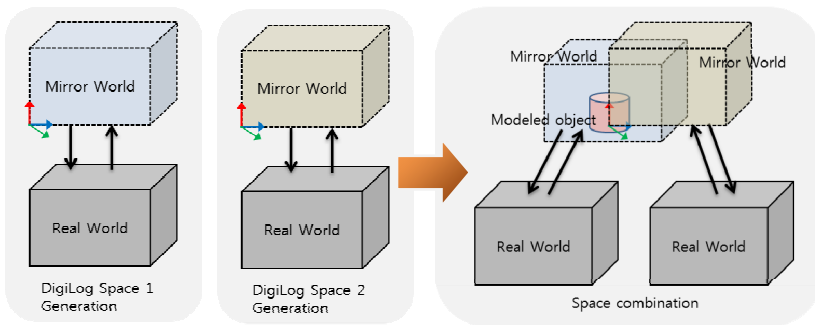
However this experience of remote collaboration is mainly possible within a spatially limited environment where users do not move. For example, in a 2-D display environment, users in fixed location can write and draw together with a remote participant or show marker based-AR information using a monitor or a large display [1-3]. In 2.5-D display environment, users can collaborate in a limited 3-D environment using a curved screen that has been installed in a fixed location [4]. On the other hand, multiple cameras or depth cameras are used in a 3-D display

environment [5-7], but tele-collaborations studied are still possible in front of a fixed camera or screen. So we need a noble method that ensures user's mobility and allows perfect 3-D interaction.

This paper suggests an augmented reality(AR) based DigiLog Space Generator. DigiLog Space Generator can generate 3-D collaboration space simply in an arbitrary and common environment. DigiLog Space is a combination of the physical world and a mirror world (digitization of the real world) [8]. The real world and mirror world are connected bidirectionally. In DigiLog Space, information is shared in real-time. DigiLog Space Generator generates these DigiLog Spaces and allows tele-collaboration through converging DigiLog Spaces. That is, the DigiLog Space Generator is composed of two parts. One is the technology of generating DigiLog Space; the other is the technology of combining several DigiLog Spaces. The overall concept and procedure is illustrated in Figure 1.

When we generate space in the real world, we also generate a mirror world. A mirror world is not physical space, but virtual space. A mirror world does not have a physical scale, and thus uses an arbitrary scale. If the scale of a mirror world and a real world are different, remote space combines with the mirror world unnaturally, because remote space and virtual information in the space are bigger or smaller than the real things. To solve this problem, we calculated the scale of generated space and saved the scale. The user combines their own space and the remote space realistically using the physical scale. We experimented and discussed the accuracy of generated space.

The remainder of this paper is organized as follows: Section 2 explains the system design of the DigiLog Space Generator. Section 3 deals with the implementation details and experimental results of in situ space modeling and combination. Finally, the conclusion and future directions of the DigiLog space generator are summarized in Section 4.



**Fig. 1.** DigiLog Space Generation and combination

## 2 DigiLog Space Generator

Figure 2 is a block diagram of a total system. It is composed of DigiLog space generation, DigiLog Space Combination, and a DigiLog Space Management Module. The DigiLog Space Generation Module reconstructs a space in real time through images from an HMD camera and user input[9]. Then it gains a 3D feature map. Next, we model an object of interest (OOI) as a reference. The modeled object has local coordinates and a scale. Then a plane is made and extruded through user input. When the system generates space using the 3D feature map, we can track the space. If space is generated, the space Id, coordinates, scale and pose are stored at DigiLog Space Management. Then the system performs a combination step for bring remote space. When a reference object is detected, the system makes a request for space sharing to DigiLog Space Management. Then the system can bring remote space at DB, and combine multiple spaces based on the reference object. Finally, the user can see the AR information and communicate with a remote partner.

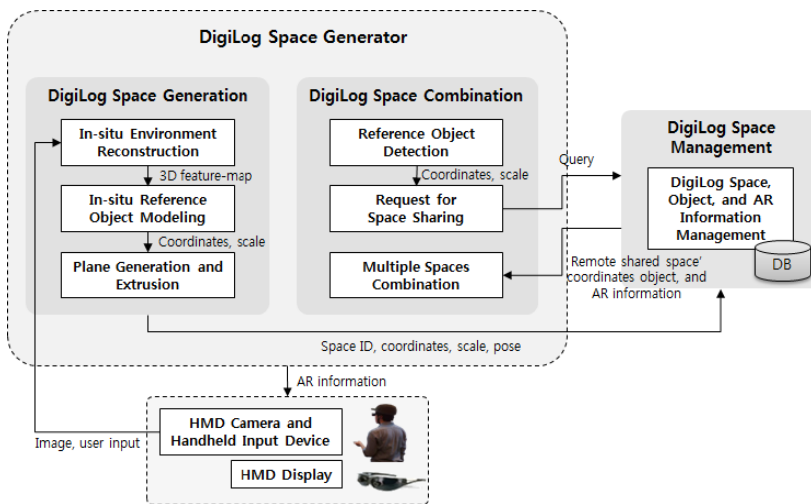


Fig. 2. Overall procedure of DigiLog Space Generator

We generate a partial space of interest according to the need to ensure the user’s mobility and share information efficiently. Because the generated space contains data for camera tracking, the HMD camera can be tracked at the space. Virtual information that we want to share can be registered in the space accurately at the HMD camera. Generally, conventional methods of space generation are conducted offline because it takes too much time to reconstruct space. These offline methods of space reconstruction are separated from the step of information sharing. Users can share information only at a fixed space that is already made. To define DigiLog space as an arbitrary environment, we propose a method of simple space modeling in real time for DigiLog Space generation. Our system captures environmental images using an HMD camera and reconstructs the environment with a 3D feature map based on the

structure-from-motion (SFM) method. A DigiLog Space is then defined by a user in the physical environment. We model only interested partial space for real time and share information selectively. Information outside the space can be hidden using an occlusion effect by generating partial space. We can define a space range that we want to share.

The modeling of an object of interest (OoI) is an important cue to combine remote space and bring augmented information. First we set the reference coordinates on the OoI, because the reference coordinates of space are needed to bring in the remote shared space and AR information. When the user selects a bottom-line reference object, reference coordinates are made at the bottom left corner of an OoI by refining coordinates. Therefore, the OoI is modeled and tracked separately based on the generated DigiLog space, with its own coordinate system. For this, multiple feature-vocabulary trees are managed for the space and the objects[10]. If we input the physical scale of the OoI, our system calculates the scale of a generated space using scale of OoI. We know the physical scale of the mirror world and save that scale at DB. We assume that the user and partner use common OoI as reference objects. The modeled, common OoI has the same feature map, scale, and coordinates, so we can combine difference spaces suitably. In other words, multiple DigiLog spaces are synchronized through connected coordinates. Therefore, sharing of AR information in real time is possible in a combined space.

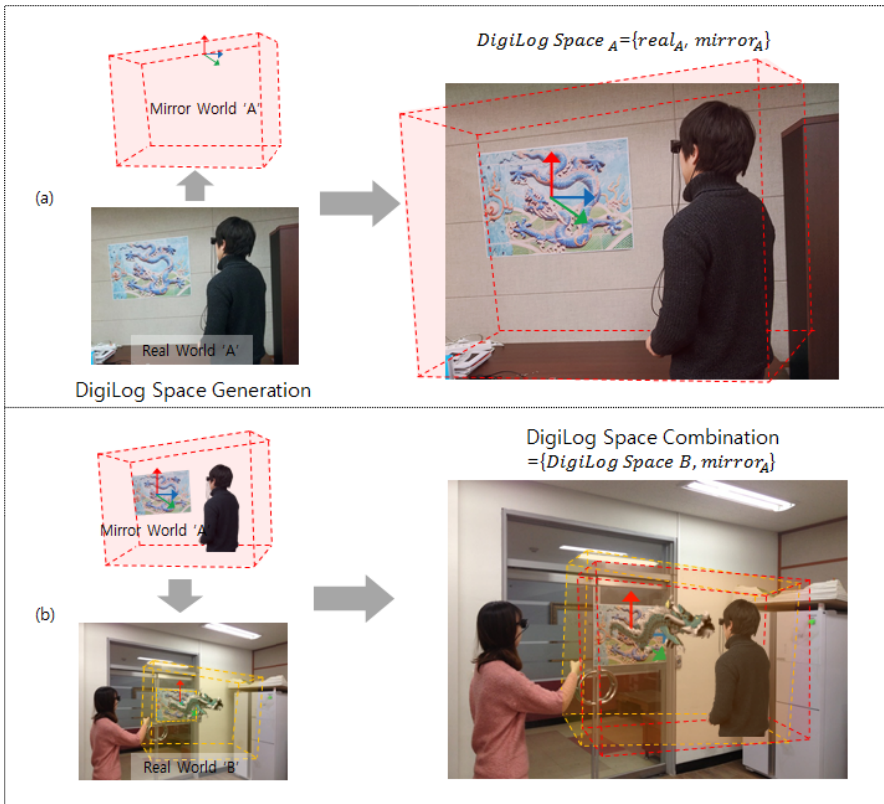
The whole scenario of the DigiLog space generator is shown in Figure 3. User “A,” wearing video see-through HMD, generates his own space and models OoI in situ with the HMD camera and a handheld input device. Input from user “A” in the real world makes the mirror world “A” at the same time. So user “A,” wearing HMD, can see the real world with the mirror world. Then user “B” combines her DigiLog space with remote space “A” based on a modeled dragon poster as a reference object created by the DigiLog Space Generator. User “B,” wearing HMD, can see both the virtual information from herself and from user “A.” Therefore user “B” can share 3D information and live communication in the 3D environment while walking. But there is a limitation in that the reference object is always seen in the user’s view.

### 3 Experiment and Implementation

To verify the performance of the DigiLog Space Generator, we measured the accuracy of generated space. We compare scale of physical space and generated space(mirror world) for measuring accuracy. Minimizing the error between of spaces is important to combine realistically. In our experiment, we set the partial space for the experiment as shown in Figure 4. The yellow box shape indicates our partial space. To make the space, the user generates coordinates at a reference object (the poster) through input. Next, the user touches four corners of the front wall to generate a plane. Finally, the plane is extruded by user interaction. Then scale of generated space is calculated and stored at DB. The user conducts the generating-space experiment 10 times.



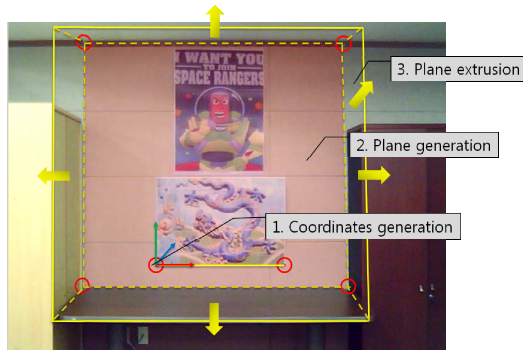
The scale of real space is 180 (x axis) x 170 (y axis) x 90 (z axis) cm. In order to generate a spatially stable space, a reference image must show all of the HMD camera view. Therefore, the user creates a space at a spot that is 4 meters away from the reference image. For stable tracking, the user only has to move the translation. The experiments were performed in a typical indoor environment. A video see-through HMD was used with 800 x 600 pixel resolution. A camera on the HMD captured 30 image frames per second. Our system was implemented using <sup>1</sup>OpenCV and <sup>2</sup>OSG.



**Fig. 3.** A whole scenario of the DigiLog Space Generator: (a) In-situ DigiLog Space Generation: 3D features extraction and space extrusion; (b) In-situ DigiLog Space Combination: a user brings remote shared space and AR information. Optionally in order to view the remote partner, a depth camera on a stand was used to detect and segment a person.

<sup>1</sup> Open computer vision library, <http://sourceforge.net/projects/opencvlibrary/>

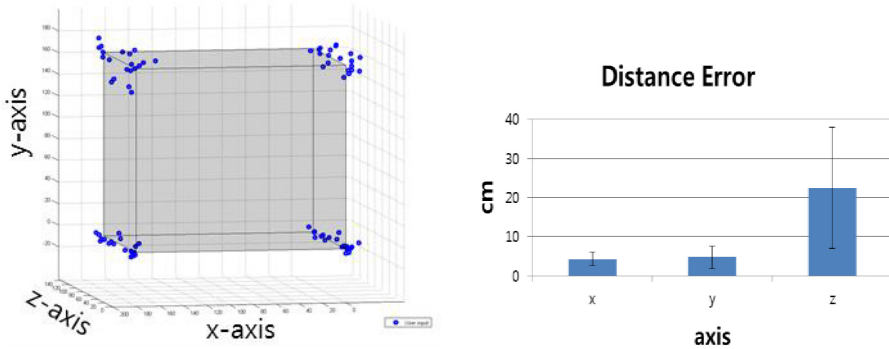
<sup>2</sup> Open scene graph library, <http://www.openscenegraph.org/projects/osg/>



**Fig. 4.** Experiment environment and steps of generating space

Figure 5(a) shows real space (a gray box) and corner points (blue points). Corner points are made by user input. Blue points mean user input. So, if the difference between corner points and box corner points is small, the generated space is similar to real space. Our generated space is pretty accurate because the blue points are crowded around the box corners. When analyzed the average of x, y, and z values, and the average errors of x and y were under 5 cm.

However, the average error of z is relatively large at 22.3cm. This points to the limitation of our system. Maybe user didn't know the point we set at the z axis because the depth value of real space is ambiguous. Also, when the user makes a plane standing at the front only, it is generated aslant because 3D points are created inaccurately. For accuracy, the user should make the plane by moving at a variety of angles.



**Fig. 5.** Experiment result: (a) Corner points of the space created by the user. The scale of the target space is 180 (x-axis) x 170 (y-axis) x 90 (z-axis) cm; (b) Distance errors between two spaces in each axis.

Fig. 6 shows an implementation of DigiLog space combination. A user combines his or her space with remote space based on modeled objects. This combined DigiLog space enables sharing 3-D information and live communication.



**Fig. 6.** A user exploits a video see-through HMD for modeling, tracking objects of interest, and sharing DigiLog space. And user can collaborate with remote partner with AR information in combined space.

## 4 Conclusion

In this paper, we introduce a novel AR-based tele-collaboration technology. Our method is more effective than a conventional, immovable video-conferencing system using cumbersome equipment and installations. We can generate interested space and combine space in real time. Our system can be used at home, or the classroom, and in other installations. For accurate combination, we use a common reference object. The reference object is important for knowing coordinates and scale of space. So our system can synchronize multiple DigiLog Spaces easily with coordinates and scale information.

Currently there are limitations for implementation. Our system does not work well in environments that have repetitive textures, or no textures. And generated space and virtual information are unstable if the OoI is occluded. An ambiguous scale of the z axis is also problematic. For future work we plan to use an RGB-D feature map, including color and depth information, for modeling and tracking. This will make generated space more stable under featureless and variant lighting environments, rather than using RGB feature points only. And if we use a depth value when generating space, the user will have a z scale, obviously.

Once our system is updated it can be applicable to many AR applications including experimental education, urban planning, military simulation, collaborative surgery, etc.

**Acknowledgements.** This work was supported by the Global Frontier R&D Program on funded by the National Research Foundation of Korea grant funded by the Korean Government (MSIP) (NRF-2010-0029751).

## References

1. Kuechler, M., Kunz, A.M.: Collboard: A remote collaboration groupware device featuring an embodiment-enriched shared workspace. In: 16th ACM International Conference on Supporting Group Work, pp. 211–214 (2010)
2. Tang, J., Marlow, J., Hoff, A., Roseway, A., Inkpen, K., Zhao, C., Cao, X.: Time Travel Proxy: Using Lightweight Video Recordings to Create Asynchronous, Interactive Meetings. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 3111–3120 (2012)
3. Barakonyi, I., Fahmy, T., Schmalstieg, D.: Remote collaboration using augmented reality video conferencing. In: Graphics Interface 2004, pp. 89–96 (2004)
4. Benko, H., Jota, R., Wilson, A.: MirageTable: Freehand Interaction on a Projected Augmented Reality Tabletop. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 199–208 (2012)
5. Schreer, O., Feldmann, I., Atzpadin, N., Eisert, P., Kauff, P., Belt, H.: 3DPresence - A system concept for multi-user and multi-party immersive 3D videoconferencing. In: CVMP, pp. 321–334 (2008)
6. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., Vertegaal, R.: TeleHuman: Effects of 3D perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 2531–2540 (2012)
7. Lehment, N., Erhardt, K., Rigoll, G.: Interface Design for an InexpensiveHands-Free Collaborative Videoconferencing System. In: ISMAR (2012)
8. Ha, T., Lee, H., Woo, W.: DigiLog Space: Real-Time Dual Space Registration and Dynamic Information Visualization for 4D+ Augmented Reality. In: ISUVR, pp. 22–25 (2012)
9. Ha, T., Woo, W.: ARwand for an Augmented World Builder. In: IEEE 3DUI (in press, 2013)
10. Kim, K., Lepetit, V., Woo, W.: Real-time interactive modeling and scalable multiple object tracking for AR. *Computers & Graphics* 36(8), 945–954 (2012)

# Onomatopoeia Expressions for Intuitive Understanding of Remote Office Situation

Kyota Higa, Masumi Ishikawa, and Toshiyuki Nomura

Information and Media Laboratories, NEC Corporation, 1753 Shimonumabe,  
Nakahara-ku, Kawasaki, Kanagawa, 211-8666, Japan  
k-higa@ah.jp.nec.com, m-ishikawa@bq.jp.nec.com,  
t-nomura@da.jp.nec.com

**Abstract.** This paper proposes a system for intuitive understanding of remote office situation using onomatopoeia expressions. Onomatopoeia (imitative word) is a word that imitates sound or movement. This system detects office events such as “conversation” or “human movement” from audio and video signals of remote office, and converts them to onomatopoeia texts. Onomatopoeia texts are superimposed on the office image, and sent to the remote office. By using onomatopoeia expressions, the office event such as “conversation” and “human movement” can be compactly expressed as just one word. Thus, people can instantly understand remote office situation without watching the video for a while. Subjective experimental results show that easiness of event understanding is statistically significantly improved by the onomatopoeia expressions compared to the video at 99% confidence level. We have developed a prototype system with two cameras and eight microphones, and then have exhibited it at ultra-realistic communications forum in Japan. In the exhibition, the concept of this system was favorably accepted by visitors.

**Keywords:** onomatopoeia, audio/video signal, remote office situation, collaborative work.

## 1 Introduction

A collaborative work between remote offices requires various communication tools such as telephone, videophone, e-mail, or online-chat. These tools frequently interrupt one’s work, which greatly decreases work productivity [1]. This problem is caused because people cannot understand remote office situation such as occurrence and level of conversation and human movement (e.g. walking or desk work), and thus cannot infer how busy the fellow worker is on the other side. Understanding remote office situation is important for a smooth communication between remote offices.

Methods have been proposed for estimating a busyness level of a remote office worker by using biological sensors [2] or PC operation records [3]. However, the scopes of these methods are limited to the users wearing sensors or using PCs.

Another approach is to watch a video of the remote office by using a surveillance camera. However, this requires people to monitor the video of remote office for a while for understanding remote office situation.

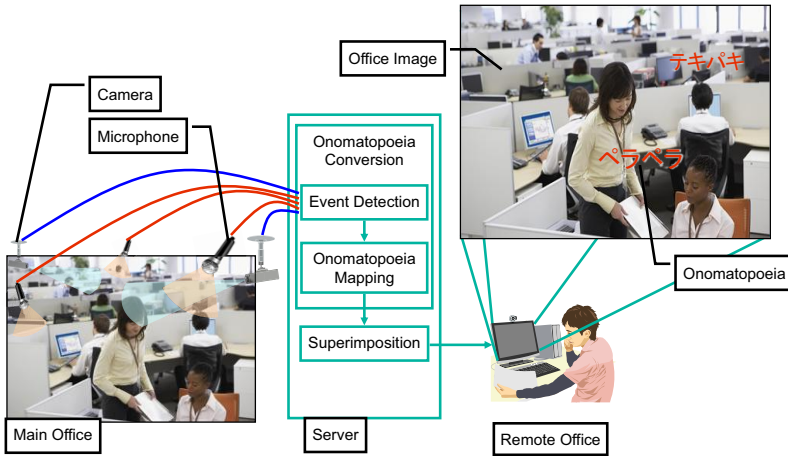


Fig. 1. System Concept

This paper proposes a system using onomatopoeia (imitative word) expressions, which enables people to instantly understand remote office situation without watching the video for a while.

## 2 System Concept

Figure 1 shows a concept of this system. The server detects office events such as “conversation” and “human movement (walking or desk work)” from audio and video signals captured by multiple microphones and cameras, and converts the office events to onomatopoeia texts. Onomatopoeia texts are superimposed on the office image, and sent to the remote office. A remote office worker can intuitively understand office situation by looking at the office image with onomatopoeia texts.

Onomatopoeia is a word that imitates the source of the sound or psychological states or bodily feelings (e.g. “whoosh” or “pitter-patter”). The onomatopoeia is often used in Japanese comics to explain details of various situations such as a sense of tension or a mental state of character.

By using onomatopoeia expressions, the office event such as “conversation” and “human movement” can be compactly expressed as just one word. Thus, people can instantly understand remote office situation without watching the video for a while. Furthermore, personal privacy is protected because details of the conversation contents are not presented to the fellow worker.

## 3 Onomatopoeia Conversion

Figure 2 shows the block diagram of onomatopoeia conversion. The audio and video signals are analyzed to detect office events, which are then mapped to onomatopoeia texts in the database.

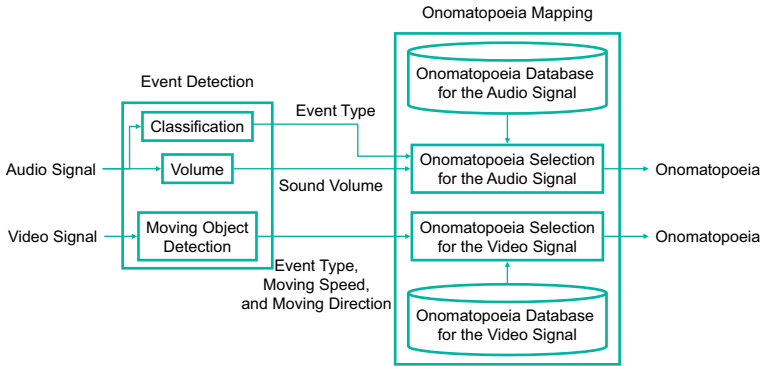


Fig. 2. Block diagram of onomatopoeia conversion

### 3.1 Onomatopoeia Conversion for Audio Signal

The audio signal is used for detecting “conversation” and its level. Onomatopoeia texts are selected based on sound volume and event type extracted from the audio signal.

The sound volume is calculated by integrating amplitude energy of audio frames. The event type is determined by classifying audio frames based on temporal change of amplitude energy and shape of frequency spectrum [4]. Each audio frame is classified into “non-conversation” (silence) or “conversation,” and then the class with highest votes in N audio frames is selected as the event type.

The sound volume and the event type are mapped to onomatopoeia text as depicted in Fig. 3. “Non-conversation” is mapped to “Shiin (describing absolute silence).” “Conversation” is mapped to “Hiso hiso (describing talk in a dim voice),” “Boso boso (describing talk in a low voice),” “Pera pera (describing fluent talk at a normal voice),” or “Gaya gaya (describing talk in a loud voice)” based on the sound volume in ascending order. These onomatopoeia texts are heuristically selected for describing “conversation.”

### 3.2 Onomatopoeia Conversion for Video Signal

The video signal is used for detecting “human movement (walking or desk work)” and its level. Onomatopoeia texts are selected based on moving direction, moving speed, and moving area of moving objects extracted from the video signal.

The moving objects are detected based on motion vectors of keypoints in the video frames. To calculate the motion vectors between adjacent frames, the keypoints are detected and tracked by Harris Corner Detector and Kanade Lucas Tomasi Tracker [5]. The keypoints with large motion vectors are grouped as a moving object.

The objects are tracked based on intersection ratio in adjacent frames to estimate the moving direction, moving speed, and moving area of moving objects. The moving

direction and speed are determined from magnitude and direction of motion vectors belonging to the objects, respectively. The moving area is a bounding rectangle including trajectory of the moving object center.

Event Type				
Non-Conversation	シーン (Shiin: describing absolute silence)			
Conversation	ヒソヒソ (Hiso hiso: describing talk in a dim voice)	ボンボン (Boso boso: describing talk in a low voice)	ペラペラ (Pera pera: describing fluent talk at a normal volume)	ギャギャ (Gaya gaya: describing talk in a group)

Fig. 3. Onomatopoeia mapping for the audio signal

Event Type		
Walking	テクテク (Teku tekku: describing walk at normal pace)	スタスタ (Suta suta: describing quick straight walk)
Desk Work	ゴソゴソ (Goso goso: describing subtle movement)	テキパキ (Teki paki: describing crisp movement)

Fig. 4. Onomatopoeia mapping for the video signal

The moving direction, moving speed, and moving area are mapped to onomatopoeia text as depicted in Fig. 4. “Walking” or “desk work” are selected based on the moving direction. “Walking” is mapped to “Teku tekku (describing walk at normal pace)” or “Suta suta (describing quick straight walk)” based on the moving speed in ascending order. “Desk work” is mapped to “Goso goso (describing subtle movement)” or “Teki paki (describing crisp movement)” based on the moving area in ascending order. These onomatopoeia texts are heuristically selected for describing “walking” and “desk work.”





Fig. 5. Examples of images with onomatopoeia texts using this experiment

## 4 Evaluation

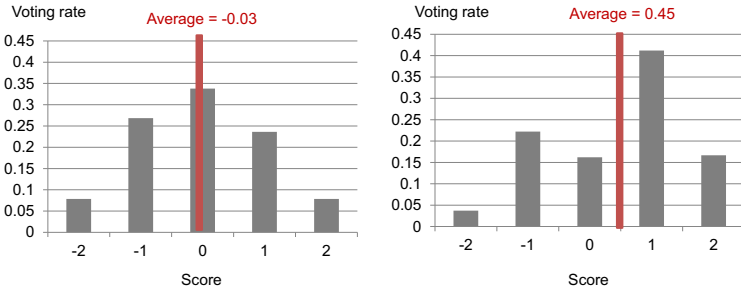
### 4.1 Experimental Conditions

We evaluated an effectiveness of the onomatopoeia expressions using 18 office events. The events consist of nine “conversation” and nine “human movement.” Each event is presented to 12 subjects in two ways: a short video (4 to 13 seconds) and an image with onomatopoeia texts extracted from the video. The subjects compared correctness and easiness of event understanding between two ways of presentation. The subjects answered a score from -2 to 2 on five-levels where the higher score means better rating of the onomatopoeia expressions. The subjects are divided into two groups. We present the video and the image in a different order with respect to each group.

The length of audio frame is 10 ms, and onomatopoeia texts for audio and video signal are superimposed on a center of the upper on the image and a position of the moving object, respectively. Figure 5 shows examples of the images with onomatopoeia texts using this experiment. Onomatopoeia texts representing “conversation” or “human movement” are superimposed on each image.

### 4.2 Results

Figure 6 shows results of subjective experiments. Figure 6 (a) shows the voting rate on correctness of event understanding, and Figure 6 (b) shows the voting rate on easiness of event understanding.



(a) Correctness of event understanding (b) Easiness of event understanding

Fig. 6. Results of subjective experiments

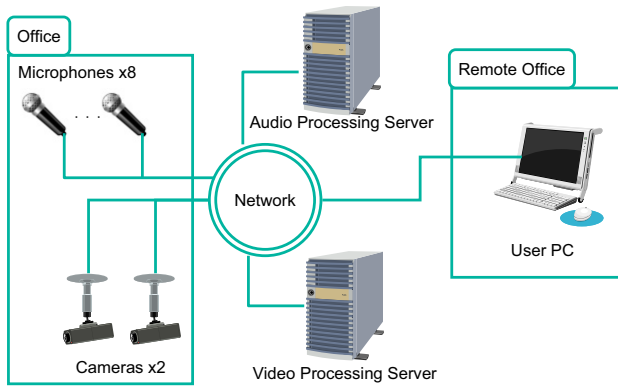


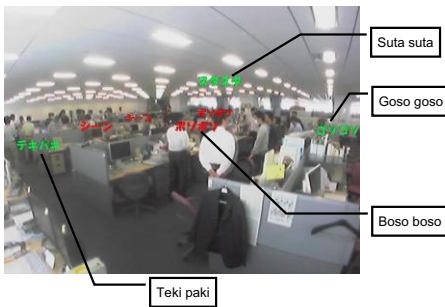
Fig. 7. System configuration of the prototype

The average score on correctness of event understanding is -0.03, which means office events can be equally understood with both ways of presenting. According to the subject comments, the onomatopoeia expressions are highly evaluated because of two advantages: (1) detecting of small events which are hard to notice by watching the video, such as slight human movements and hush conversations, and (2) correctly finding of a place where each event occurs. On the other hand, two disadvantages of onomatopoeia expressions are extracted: (1) misunderstanding of office events when onomatopoeia texts are superimposed on wrong positions, especially conversation, and (2) strange feeling from mismatch between office events and onomatopoeia texts.

The average score on easiness of event understanding is 0.45, which indicates office events can be more quickly understood by the images with onomatopoeia texts than the videos. The difference between both ways of presenting is statistically significant at 99% confidence level according to the t-test. The subjects favorably accept that the onomatopoeia expressions enable to instantly understand “walking” which requires a few seconds to see whole of the event.

**Table 1.** Specification of equipment for prototype system

Audio/Video Processing Server	HP xw6600 Workstation, Intel® Xeon® CPU E5450 @ 3.00GHz, 3.25 GB RAM, Windows XP Professional Service Pack 3
Microphone	Sony ECM-C10
Network Camera	Axis 2100
User PC	Windows XP Professional Service Pack 3, Intel® Core™2 Duo CPU E6850 @ 3.00GHz, 1.96 GB RAM



**Fig. 8.** Office image with onomatopoeia texts



**Fig. 9.** Scene of exhibition

The onomatopoeia expressions are effective for intuitive understanding of office situation while the correctness of event understanding is comparable. To improve the onomatopoeia expressions, the following functions should be implemented: (1) detection of appropriate superimposing positions to prevent misunderstanding of office events and (2) selection of onomatopoeia texts suitable for the office events depending on user’s preference.

## 5 Prototype System

We developed a prototype of the proposed system shown in Fig. 7. This system uses eight microphones and two cameras to monitor approximately 16 people. The specification of the system equipment is shown in Table 1. Onomatopoeia texts for audio and video signal are superimposed on a position of microphone in the office image and the moving object, respectively.

Figure 8 shows an office image with onomatopoeia texts of a scene that people get together for a short conversation. Red texts represent “conversation” and green texts represent “human movement.” Onomatopoeia text such as “Suta suta (its meaning is shown in Fig. 4)” or “Goso goso” for describing human movements, and “Boso boso” for describing conversation are superimposed. The remote office worker can instantaneously understand that people are gathering and having a talk.

We have exhibited this system at ultra-realistic communications forum in Japan. As shown in Fig. 9, the system presented the events in the conference room. In the exhibition, the concept of this system was favorably accepted by visitors. We also got comments which recommend applying the onomatopoeia expressions to entertainment applications.

## 6 Conclusion

We have proposed a system for intuitive understanding of remote office situation using onomatopoeia expressions. This system detects office events such as "conversation" or "human movement" from audio and video signals of remote office, and converts them to onomatopoeia texts. Onomatopoeia texts are superimposed on the office image, and sent to the remote office. By using onomatopoeia expressions, the office events can be compactly expressed as just one word. Thus, people can instantly understand remote office situation without watching the video for a while. Subjective experimental results showed that easiness of event understanding is statistically significantly improved by the onomatopoeia expressions compared to the video at 99% confidence level. We have developed a prototype system with two cameras and eight microphones, and then have exhibited it at ultra-realistic communications forum in Japan. In the exhibition, the concept of this system was favorably accepted by visitors.

**Acknowledgments.** This work is partly supported by National Institute of Information and Communications Technology (NICT), Japan.

## References

1. Mark., G., Gonzalez., V.M., Harris, J.: No Task Left Behind? Examining the Nature of Fragmented Work. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pp. 321–330 (2005)
2. Chen, D., Hart, J., Vertegaal, R.: Towards a Physiological Model of User Interruptability. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 439–451. Springer, Heidelberg (2007)
3. Tanaka, T., Fujita, K.: Interaction Mediate Agent Based on User Interruptibility Estimation. In: Human-Computer Interaction International, pp. 152–160 (2011)
4. Recommendation ITU-T G.720.1: Generic Sound Activity Detector. ITU-T (2010)
5. Shi., J., Tomasi., C.: Good Features to Track. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994)

# Enhancing Social Presence in Augmented Reality-Based Telecommunication System

Jea In Kim, Taejin Ha, Woontack Woo, and Chung-Kon Shi

Graduate School of Culture Technology, KAIST, Republic of Korea  
{jeainkim86, taejinha, woo, chungkon}@kaist.ac.kr

**Abstract.** The main contribution of this paper is to examine the new method of augmented reality from a telecommunication point of view. Then, we tried to present the fact that the concept of social presence is an important cue for developing telecommunication system based on augmented reality technology. The evaluation was conducted with 32 participants. According to the questionnaires results, the augmented reality based telecommunication system was better than 2 dimensional based display telecommunication system. To develop our concept, we should closely analyze communication patterns and improve our augmented reality based communication system.

**Keywords:** Telecommunication, Augmented Reality, Social Presence.

## 1 Introduction

Between remotely located people, Telecommunication systems have consistently developed from mobile device to a networked virtual environment such as computer-supported cooperative work (CSCW). In this context, augmented reality (AR) systems supplement the real world with virtual objects that appear to coexist in the same space as the real world [1] and create a face-to-face type of environment. However, the situation of optimal communication is the case of face-to-face communication in practice because cues of natural nonverbal communication exist [2][3]. Thus, some researchers presented social presence as one of the key features of the telecommunication that is focused on users' social psychological properties [4-6]. Especially, understanding of social presence's factors is important for supporting the high level of social presence in the telecommunication.

## 2 Related Work

### 2.1 Social Presence in Telecommunication

Many researchers studied ways of keeping a social presence with remote person in telecommunication. Hauber and colleagues explored ways to combine the video of a remote person to best emulate face-to-face cooperation with a shared tablet display [7].

They compared the values of social presence of a standard non-spatial two-dimensional (2D) interface with four kinds of conditions, not AR. Meanwhile, the system was made for enhancing the social presence of telecommunication [8]. Although researchers proposed an AR approach, their system did not have proper factors of AR but merely expanded the closed 2D display. The TeleHuman three-dimensional (3D) videoconferencing system supports 360-degrees motion parallax as the viewer moves around the cylinder and stereoscopic 3D display of the remote person with a high sense of social presence [9]. However, one-way telecommunication system could not lead to real communication between both of the users. Besides, the motion cannot solidify enough of the factor of a high sense of social presence for effective communication. Through related works, our vision was that we considered not only technology-centered thought, but also user-centered thought. Therefore, we made an effort to identify the factors of maximizing social presence for understanding the full potential of AR to telecommunication experience.

## **2.2 Augmented Reality**

Augmented reality was defined as “a continuum of real-to-virtual environments, in which augmented reality is one part of the general area of mixed reality” [10]. Azuma specified augmented reality as augmenting the real world environment with virtual information by improving people’s senses and skills [1]. He also mentioned three common characteristics of augmented reality scenes; combination of the real and virtual, interactive in real time and having the scenes registered in 3D. Although these explanations were greatly obvious, they did not get enough attributes for defining effective communicating tool as AR-based telecommunication. Thus, the concept of augmented reality must be redefined maximizing the effect of social presence effect on telecommunication.

## **3 The Three Factors of Social Presence**

The purpose of this study is to find the proper factors of social presence to the experience of AR-based telecommunication, make a prototype based on found the factors found and conduct empirical evaluation. That is, we make an effort to identify factors that enhance a sense of social presence for understanding the full potential of AR in the telecommunication experience. Therefore, this paper presents the new characteristics of AR considering the concept of social presence based on the three common characteristics of Azuma’s study.

### **3.1 A Sense of Being Together**

AR technology supports the combination of the real environment and a virtual environment [1]. Towell and Towell presented an understanding of the contribution to presence through social interaction in other virtual environments [11]. They found an important factor in the user’s experience of being with others is a sense of social

presence. In AR-based telecommunication, users can feel a sense of being with another person who is in another virtual environment.

### **3.2 A Sense of Spatial Co-presence**

AR has scenes registered in 3D [1]. When a person and a remote person share an AR's scene, they can feel a spatial co-presence, which is one of the factors associated with a sense of social presence [12]. Meanwhile, conducting work with common interesting thing leads to a feeling of high cohesion [13]. Therefore, AR-based telecommunication needs common interesting things to enhance social presence.

### **3.3 A Sense of Psychological Involvement as Mediated Social Presence**

AR includes properties of interaction in real time [1]. This interaction means psychological involvement as well as physical communication. Biocca et al. focused on mediated social presence, which means how much users involve themselves psychologically [14]. Psychological involvement is one of the core factors to obtain a feeling of social presence.

## **4 Implementation**

Fig. 2. shows a block diagram of our system, which captures real-space images from an HMD camera, conducts camera pose tracking, and then registers with virtual objects. Specifically, in the offline camera tracking data was obtained via the 3D geometrical structure of a real-world environment. Then, a local reference coordinate was allocated on the planar object (e.g., a picture attached to the wall) in front of a user [15].

In the implementation, the virtual objects are the textured human and some objects in the other space that are augmented based on the local reference coordinate. To do this, the system in the other space detects and segments the foreground objects from the learned environment by using a red-green-blue depth camera (RGB-D camera), and then it transmits the foreground objects to the in-situ user through wired communication.

The hardware used in the implementation included a bi-ocular video see-through HMD [16] with  $800 \times 600$  pixel resolution. A camera attached to the HMD captured 30 images per second with  $640 \times 480$  pixel resolution. A RGB-D camera [17] in front of a user captured 30 RGB images per second with  $640 \times 480$  pixels and 30 gray-scale depth images per second with  $320 \times 240$  pixels for objects within 1.2 to 3.5m. The computer used was equipped with an independent graphics card and a core of the i7 central processing unit (CPU).

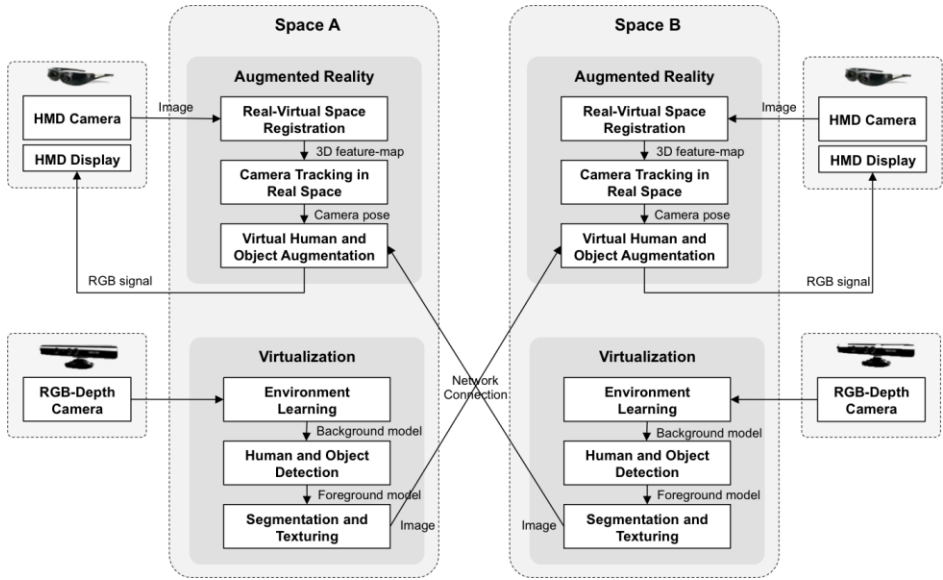


Fig. 1. A block diagram of our AR-based telecommunication system

## 5 Empirical Evaluation

We designed an experiment to evaluate the suitability of three factors based on the concept of AR and social presence. An experiment focused on how users can feel a sense of social presence when they communicate spontaneously with the common interesting thing. Thus, participants conducted a task-based experiment on a 2D display-based telecommunication system [18] and the AR-based telecommunication system we developed, respectively.

### 5.1 Participants

The study included 32 participants. The average of age was 24.3 years old, ranging from 19 to 34. Every participant had experienced the telecommunication system at least once. Gender was balanced in the ratio between men and women (i.e., 16 men and 16 women). That is, there is no gender constraint in measuring social presence. Two participants conducted the task as a team.

### 5.2 Task

A pair of participants were assigned the roles of manipulator and facilitator. When they decided on their role, the manipulator sat in a chair with a puzzle plate that could accommodate the puzzle pieces. The facilitator sat in the chair with an image of the completed puzzle. These participants stayed separated in different locations. The pair



of participants put the puzzle together with free verbal and nonverbal communication. In the process of putting the puzzle together, the facilitator could help the manipulator because the facilitator had more information about the completed image. The manipulator could not look at the facilitator's completed image. The participants experienced two kinds of telecommunication (i.e. AR and 2D video). For each team, the test schedules were separated by more than two weeks to prevent any learning effects.

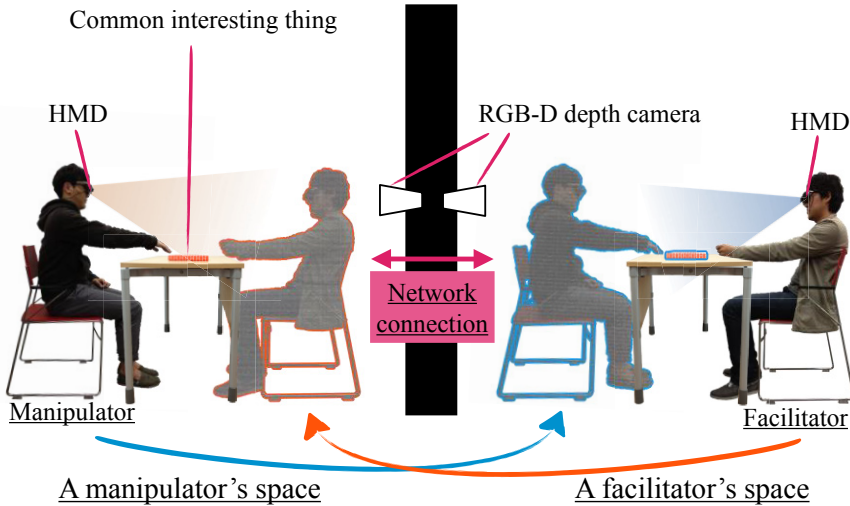


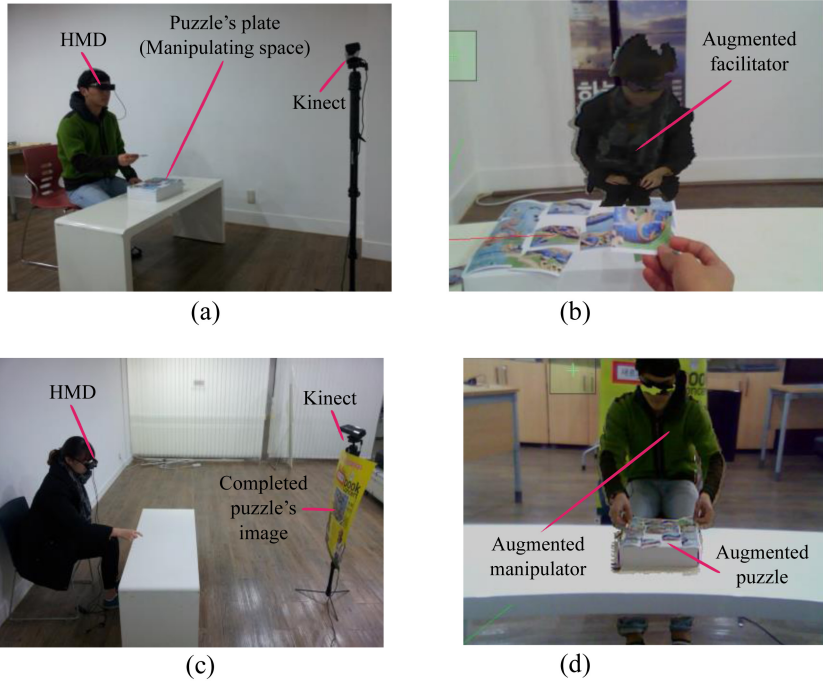
Fig. 2. The teleconferencing environment

### 5.3 Experiment Design

A within-subjects' design was used. At first, the participants were given an explanation of our task-based experiment. They were given time to practice the systems (i.e. the 2D display-based telecommunication system and the AR-based telecommunication system) for completing a puzzle (i.e. the common interesting thing). Then, the participants answered questionnaires after experiencing the system, respectively. The degree of social presence was measured according to the results of a comparative analysis.

### 5.4 Questionnaire Construction

To evaluate the degree of social presence, the participants answered seven-point Likert scale questionnaires after each task. In the questionnaire, 1 means "strongly disagree" and 7 means "strongly agree". Three kinds of social presence factors are addressed in a total of 12 items; a sense of being together is related to one item [11], a sense of spatial co-presence is related to one item [12] and a sense of psychological involvement is related to 10 items [14].

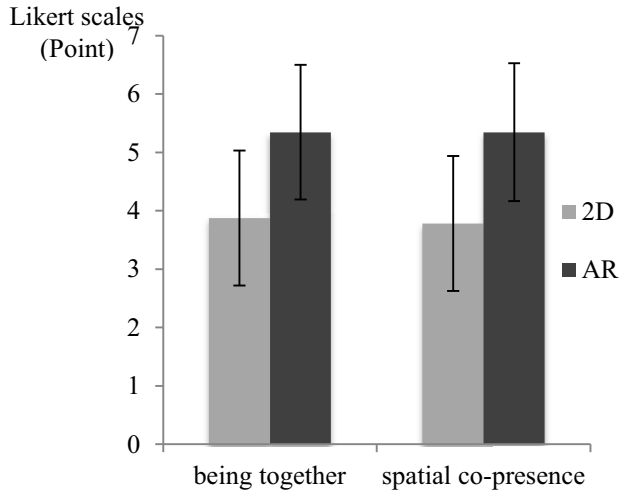


**Fig. 3.** The actual experimental environment of the AR-based telecommunication: Each user can see his or her virtual remote partner through the video see-through HMD and thus communicate. (a) Manipulator stay with puzzle’s plate, (b) Manipulator’s view through HMD, (c) Facilitator stay with complete puzzle’s image, (d) Facilitator’s view through HMD.

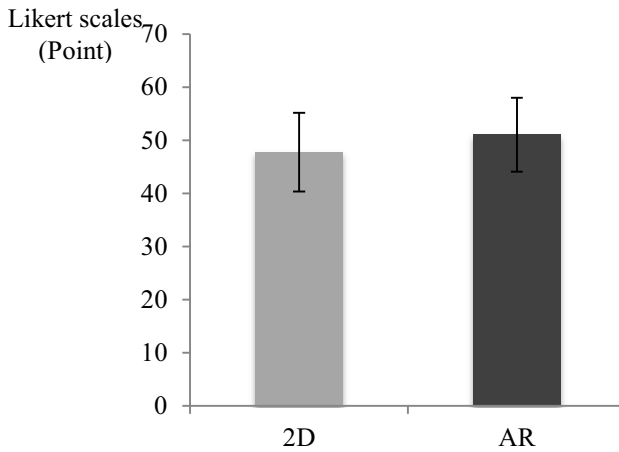
### 5.5 Results

In dealing with the questionnaire results, this paper took a survey of 32 participants (i.e., 16 pairs) and analyzed each survey item’s reliability. For the comparative analysis, the questionnaire results were analyzed using a within-subjects analysis of variance (i.e., paired samples t-test), evaluated at an alpha level of .05.

These results indicated that the participants felt a significantly greater sense of being together ( $p < 0.05$ ) and spatial co-presence ( $p < 0.05$ ) with AR-based telecommunication. However, there was no significant difference in a total of 10 questions related to a sense of psychological involvement as mediated social presence ( $p = 0.113$ ).



**Fig. 4.** The graph of questionnaire results about a sense of being together and spatial co-presence



**Fig. 5.** The graph of questionnaire results about a sense of psychological involvement as mediated social presence

**5.6 Discussion**

According to the questionnaire result, an AR-based telecommunication system could enable participants to feel a sense of being together and spatial co-presence. That is, the telecommunication with AR technology with HMD is a better way to communicate together like face-to-face than telecommunication with a 2D display. However we cannot sure that the suggested factors are the best way to conduct AR-based telecommunication. For a better telecommunicating experience, we would clarify which of our factors are essential.

In the case of the factor of psychological involvement, we should analyze more detailed information in our experimental environment. The participant actually communicated together with each other in verbal and non-verbal ways. Therefore, we can explain the results precisely when we analyze the participants' communication patterns.

Finally, our AR-based telecommunication system requires adjustments to address technical issues. In conducting the usability test, we found some technical problems such as turning off and being suddenly unable to look at augmented people.

## 6 Conclusion

The purpose of the current study was to identify the proper factors of social presence in experiencing AR-based telecommunication and to confirm the effectiveness by conducting an empirical evaluation. We proposed three AR-based telecommunication factors based on previous concepts of social presence. The evaluation was conducted with 32 participants. According to the questionnaires results, the AR-based telecommunication system was better than 2D display-based telecommunication system in terms of feeling a sense of being together and spatial co-presence. However, there was no significant difference in the result for psychological involvement.

To develop our concept, future research should closely analyze communication patterns. Furthermore, we should improve our AR-based telecommunication system to address technical and visual issues.

The present study examined a new method of augmented reality from a telecommunication point of view. Then, we tried to present the fact that the concept of social presence is an important cue for developing telecommunication system based on AR technology. Additionally, conducting a well-ordered empirical evaluation would be helpful in designing future telecommunication system.

## References

1. Azuma, R.T.: A Survey of Augmented Reality, Presence: Teleoperators and Virtual Environments, pp. 355–385 (August 1997)
2. Walther, J.B., Burgoon, J.K.: Relational Communication in Computer-mediated Interaction. *Human Communication Research* 19(1), 50–88 (1992)
3. Walther, J.B.: Computer-mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research* 23(1), 3–43 (1996)
4. Short, J.A., Williams, E., Christie, B.: *The Social Psychology of Telecommunication*. John Wiley & Sons, Ltd., London (1976)
5. Tu, C.-H.: The Measurement of Social Presence in an Online Learning Environment. *International Journal on E-learning* 1(2), 34–45 (1992)
6. Lee, M.N., Nass, C.: Designing Social Presence of Social Actors in Human Computer Interaction. In: *Proceedings of the the SIGCHI Conference on Human Factors in Computing System*, pp. 289–296 (April 2003)

7. Hauber, J., Regenbrecht, H., Billinghurst, M., Cockburn, A.: Spatiality in Videoconferencing: Trade-offs between Efficiency and Social Presence. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 413–422 (November 2006)
8. Almeida, I.S., Oikawa, M.A., Carres, J.P., Miyajaki, J., Kato, H., Billinghurst, M.: AR-based Video Mediated Communication: A Social Presence Enhancing Experience. In: Virtual and Augmented Reality, pp. 125–130 (May 2012)
9. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., Vertegaal, R.: TeleHuman: Effects of 3d Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2541–2540 (May 2012)
10. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented Reality: A Class of Displays on the Reality-virtuality Continuum. In: Telem manipulator and Telepresence Technologies, pp. 282–292 (November 1994)
11. Towell, J.F., Towell, E.: Presence in Text-Based Networked Virtual Environments of MUDS, pp. 590–595. MIT Press (September 1997)
12. Mason, R.: Using Communications Media in Open and Flexible Learning. Kogan Page, London (1994)
13. Wenger, E.: Communities of Practice. Cambridge University Press (1998)
14. Biocca, F., Harms, C., Gregg, J.: The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. Paper presented at the Presence 2001 Conference (2001)
15. Kim, K., Lepetit, V., Woo, W.: Real-Time Interactive Modeling and Scalable Multiple Object Tracking for AR. Computers and Graphics 36(8), 945–954 (2012)
16. Wrap™ 920AR, <http://www.vuzix.com>
17. MS Kinect sensor, <http://www.microsoft.com/en-us/kinectforwindows/>
18. Skype, <http://www.skype.com>

# How Fiction Informed the Development of Telepresence and Teleoperation An Historical Perspective

Gordon M. Mair

Transparent Telepresence Research Group,  
Department of Design, Manufacture, and Engineering Management,  
University of Strathclyde, 75 Montrose Street, Glasgow, G1 1XJ, Scotland, UK  
g.m.mair@strath.ac.uk

**Abstract.** This paper shows that many telepresence and teleoperation innovations and patents actually had their precursors in fiction and that this led the way for technological developments. This suggests justification for those companies that have invested, or are considering investing, in funding science fiction writers to provide future scenarios for their respective products and industries. The research leading to this conclusion has involved a search of patents, technical and scientific publications, and fictional works. The paper is mainly concerned with telepresence and teleoperation but aspects of virtual reality are included where the technological and literary concepts are relevant.

**Keywords:** Virtual reality, presence, teleoperation, science-fiction, telepresence history.

## 1 Introduction

For over a century fiction has speculated on a number of scenarios involving what we now call ‘telepresence’. Throughout the same period technologies have developed that allowed the practical creation of aspects of telepresence. It is normal for authors of fiction, and particularly science fiction, to exploit existing and emerging scientific developments as a basis for their stories. However this paper suggests that in some instances fiction not only preceded but also inspired and led scientific and technological thinking in important aspects of telepresence and teleoperation. This is significant as the evidence adds credibility to the notion that close attention to science fiction by scientists, engineers, and commercial entities can provide vital inspiration for real world innovation and commercial development. For example a well-known mobile phone company once commissioned award winning science-fiction authors to write a book of short stories on the future of mobile communication [1]. This provided a source of possible product and application ideas as well as speculation on the social and cultural impact of developments.

The fiction of most relevance is of the type where the means for the achievement of the technology is described, sometimes called hard science fiction. The focus is

therefore on technological manifestations of telepresence and teleoperation and the following paper examines how both fact and fiction created tributaries that led to the development of today's telepresence systems.

We can consider this topic using two time periods. Seminal fictional ideas and technical inventions that would lead to telepresence occurred very noticeably during the period 1876 to 1938. However throughout these years the means whereby the fictional ideas could be manifested were very vague. For telepresence itself much of the directly relevant activity in both fact and fiction occurred from around the middle of the 20th Century, including what has been arguably called the Golden Age of Science Fiction between 1939 and 1959, to the present time of writing in 2013.

## 2 The Tributary Years 1876 to 1938

### 2.1 Telephonoscopes and Nipkow Disks

Significant audio transmission development began when Scottish born Alexander Graham Bell applied for his patent on the telephone "An Improvement in Telegraphy" in 1876 [2]. On the 24<sup>th</sup> of December 1877 Thomas Edison filed his patent application for an "Improvement in phonograph or speaking machines" [3]. Since we are concerned here with telepresence it is interesting to see that the Scientific American of the 22<sup>nd</sup> of December 1877, two days *before* Edison's filing, carried an article, including an illustration of his device and stating; "*It is already possible by ingenious optical contrivances to throw stereoscopic photographs of people on screens in full view of an audience. Add the talking phonograph to counterfeit their voices, and it would be difficult to carry the illusion of real presence much further*" [4].

These factual events quickly fired imagination and, considering that moving pictures were beginning to appear from a number of sources, Bell's and Edison's inventions inspired 'Punch' to publish an imaginative sketch on December the 9<sup>th</sup> 1878 of a "Telephonoscope" in which a mother and father converse live with their daughter thousands of miles away using what appears to be a super widescreen television and audio system more advanced than today's telepresence systems.

With regard to the beginnings of transmitted video, by 1884 existing knowledge of electricity and photoconductivity was put to use by a German university student Paul Gottlieb Nipkow who invented the Scanning Disk when only 20 years old. Although a working system does not appear to have been built, this disk allowed the possibility of live transmission of very basic, low resolution, flickering images over a distance using an electro-mechanical system [5]. Not long after this in 1888 the story "In the Year 2889" by Jules Verne and his son Michel Verne was published [6]. In the story a very basic description is given of an audio-visual system similar to what we would call today a videophone.

### 2.2 The Vision of Wells and Contemporaries

"The Remarkable Case of Davidson's Eyes" [7] first published in 1895 saw H.G. Wells present the concept of someone sensing they were at a location remote from their physical body due, albeit very indirectly, to a type of technological mediation. Concerning Davidson, Wells says "In some unaccountable way, while he moved

hither and thither in London, his sight moved hither and thither in a manner that corresponded about this distant island". The island was located in the Pacific Ocean on the other side of the world from London. This is perhaps the first description of immersive telepresence. "The Crystal Egg" [8] followed two years later in which Wells describes a system that allows a viewer on Earth to observe moving images of Mars – a foretelling of today's unmanned Martian rovers. Of course, just as with today's robotic planetary explorers, a means of controlling the remote cameras would be necessary. However the first indication of how this could be achieved in practical terms came with a public demonstration a year later by Nikola Tesla in 1898 when he used "coded pulses via Hertzian waves" to radio control a model submersible boat in Madison Square Garden [9].

Returning to the theme of visual sensing, transmission, and display; our modern word *television* was first used by Par M. Constantin Perskyi in a paper titled "Television Au Moyen De L'electricite". This was presented to the 1<sup>st</sup> International Electricity Congress at the World Fair in Paris on August 25<sup>th</sup> 1900 [10]. Then in June 1908 Nature published a letter sent by English physicist and inventor Shelford Bidwell which was titled "Telegraphic Photography and Electric Vision" [11]. Two weeks later a response by the Scottish electrical engineer Alan Archibald Campbell-Swinton was published and titled "Distant Electric Vision" [12]. It described how cathode ray tubes could be used for both acquiring an image at the transmission end and displaying an image at the receiving end for distant electric vision. The letter said the following; "Referring to Mr. Shelford Bidwell's illuminating communication on this subject published in Nature of June 4 1908, may I point out that though, as stated by Mr. Bidwell, it is wildly impracticable to effect even 160,000 synchronised operations per second by ordinary mechanical means, this part, of the problem of obtaining distant electric vision can probably be solved by the employment of two beams of cathode rays (one at the transmitting and one at the receiving station) synchronously deflected by the varying fields of two electromagnets placed at right angles to one another and energised by two alternating electric currents of widely different frequencies, so that the moving extremities of the two beams are caused to sweep synchronously over the whole of the required surfaces within the one-tenth of a second necessary to take advantage of visual persistence". This is generally thought to have been the earliest description of how television could be obtained electronically at both transmission and reception by non electromechanical means.

However despite Campbell-Swinton's description the actual transmission of visual images did not begin until the 1920s. This occurred with the pioneering television work of Scottish inventor John Logie Baird using the electromechanical Nipkow Disk system. Baird and Day filed a patent application on the 25<sup>th</sup> of July 1923 for "A system of transmitting views portraits and scenes by telegraphy or wireless telegraphy" and the patent was awarded the following year [13], forty years after Nipkow's patent. Baird's first demonstration of moving images was in what was then Selfridge's department store in London in 1925 and this was followed by a public demonstration to members of the Royal Institution in 1926. Like other inventors and scientists of his day Baird was known to have read H.G. Wells and therefore likely that he had read Davidson's Eyes and The Crystal Egg written almost thirty years earlier.



On December the 29<sup>th</sup> 1923 the Russian-American Vladimir K. Zworykin filed his patent application for a “Television System” although it was not granted until 1938 [14]. On January the 7<sup>th</sup> 1927 the American Philo Farnsworth filed for patents for both a “Television System” [15] and a “Television Receiving System” [16] and was granted patents on them on the 26<sup>th</sup> of August 1930. Despite Baird achieving many television firsts with his electromechanical system, e.g. first transatlantic television signal in 1928 and first live television transmission of the Epsom Derby in 1931, it was electronic television that was eventually adopted worldwide.

### 2.3 A Brave New World

In the realm of fiction we now have, in 1932, Aldous Huxley’s utopian-dystopian fantasy of the future “Brave New World” [17]. Of particular relevance to the topic of immersive telepresence is his description of the “Feelies”. The following quotes occur when the two protagonists visit the “Feelie” theatre. “The scent organ was playing a delightfully refreshing Herbal Cappriccio .....”. Then later “The house lights went down; fiery letters stood out solid as though self-supported in the darkness Three Weeks In A Helicopter. An All Super-Singing Synthetic-Talking, Coloured, Stereoscopic Feely. With Synchronised Scent-Organ Accompaniment”. There follows a description of a simple but very sensual story in which, for example, “... the stereoscopic lips came together again, and once more the facial erogenous zones of the six thousand spectators in the Alhambra tingled with almost intolerable galvanic pleasure”. The concept of stereoscopic images had been around since Wheatstone in 1838 [18] and they had been extremely popular in the Victorian era but Huxley’s fictional use of olfaction and haptics as part of cinema entertainment is novel. This concept was to be reintroduced as a practical possibility 18 years later by Mort Heilig [19].

As we come to the end of this first period we see the concept of full telepresence being introduced through the medium of an anthropomorphic telepresence robot. This occurs within a short story “The Robot and the Lady” written by Manly Wade Wellman and published in Thrilling Wonder Stories in 1938 [20]. Here the protagonist uses a robot he has created as a surrogate to go on a date for him with a girl he has never met before. “My prize robot, tall, dashing would speak and act for me... I turned to where, on my desk, I had set up my controls. To my ears I clamped receivers, upon my eyes I bound the goggle like televisions that would coincide my viewpoint with that of the robot. A transmitter would place my voice upon those sculptured lips.” then “My toe pressed a switch. At once, my vision-point changed. I seemed to sit on the bed’s edge, gazing through the robot’s pupils. I touched the keys, and rose to my - but that was an illusion, born of years of such experiments. I remained silent, but the robot rose. I moved it across the floor, closed its fingers around the doorknob, and set it out into the hall. My hearing, vision and awareness went along with that excellent imitation of a young Adonis...”. It transpires that the girl also sends a surrogate robot, but in the end, in a manner similar to what might happen today, they meet face to face in their real bodies and fall in love.

### 3 The Confluence Years 1939 to 2013

#### 3.1 Heinlein and Waldo

“Waldo” was a science-fiction story by Robert Heinlein writing under the pseudonym Anson MacDonald and published in *Astounding Science Fiction* in August 1942 [21]. The story title gave us a word that came to be used as an early generic term for all remotely controlled manipulating arms. I can find no evidence for scientific precursors to his concept of these telemanipulators, although they are loosely implied in earlier ‘robot’ stories like Wellman’s mentioned above.

I suggest this is an atypical instance of a science-fiction writer coming up with a practical idea for technology before it was considered in any detail by scientists or technologists. The subject of the story is a highly intelligent individual called Waldo F. Jones who suffers from myaesthesia gravis. This has caused his muscles to be so weak he cannot comfortably live on Earth so he lives in the weightlessness of space in an orbiting satellite he has commissioned. He also builds master–slave manipulators, called primary and secondary “waldos”. The secondaries are of various sizes some of which have hands “the size of a man’s body” that can lift massive steel plates, and others are at the microscopic level for neurosurgery since he also uses small waldos to create smaller ones and so on. This miniaturisation idea however was not new as Raymond Z. Gallun had earlier introduced the concept of microrobots building smaller microrobots that then built smaller microrobots down to the atomic level [22].

At one point in the story two visitors are going to see Waldo and one of them says that the satellite “...must be all of twenty-five thousand miles up”. When they arrived “Waldo F. Jones seemed to be floating in this air at the centre of a spherical room. The appearance was caused by the fact that he was indeed floating in air. His house lay in a free orbit, with a period of just over twenty-four hours.” Heinlein appears to be describing here a satellite in synchronous orbit from which he can communicate by radio with Earth since a geosynchronous orbit occurs at 22,236 miles. This is three years before the Arthur C. Clarke letter to “Wireless World” in February 1945 proposing geosynchronous communication satellites [23]. However it is possible Heinlein could have read of the idea in an English translation of a large part of Herman Potocnik’s book “The Problem of Space Travel – The Rocket Motor”. In the book the concept of a space station is described and the height of a geostationary orbit calculated. This translation appeared as a series in the American “Science Wonder Stories” throughout July, August, and September 1929 [24].

Where did Heinlein get the idea of ‘waldos’? It seems that, although it is not an actual invention since he did not describe it in sufficient detail for it to be patentable, this is a real inventive concept not copied from existing technology. Since in the fictional story he titles his patent “Synchronous Reduplicating Pantograph” I carried out a patent search for real patents from before 1942 using similar type kinematic configurations. The only ones I found related to pantographs were primarily for engraving e.g [25]. However there are a number of patent applications relating to position controllers for paint spraying guns e.g. [26, 27] and an electro-mechanical ‘Handle Control System’ [28] for control of gun movements. Although indirectly related to some aspects of the waldos none of these are close to Heinlein’s idea and even had he been aware of these patents his waldos remain apparently original in concept for the time.

Subsequent to “Waldo” however, and due to the work in the emerging nuclear industry, remote handling became much more of a practical necessity. From the mid 1940s the need to remotely handle radioactive materials in the nuclear industry led to the development of mechanical telemanipulators in places like the Argonne National Laboratories in the USA. These devices had bilateral control and by the 1950s electro-mechanical telemanipulators were in operation incorporating servo-control and force feedback [29]. This meant that the link between the user and the end effector of the manipulator could be a direct mechanical, electrical, or hydraulic connection. Goertz, one of the inventors in the previously mentioned Handle Control System, is prominent in the development of this equipment. His name appears on many relevant patents originally filed between the early 1940s and early 1960s. The early patents are related to control methods e.g. electro-mechanical [30] and electro-hydraulic [31], with the later ones relating to full handling systems e.g. [32].

A significant first occurred in 1943, this was the filing of a patent for a head mounted display [33]. This was by Henry J. De N. McCollum for a “Stereoscopic Television Apparatus” incorporating two miniature cathode ray tubes attached to the front of a pair of lenses held in a spectacle frame.

### 3.2 Virtual Worlds and Distant Worlds

As the second half of the twentieth century begins we find that much of the technology required at the ‘home site’ of a telepresence system is common to that of ‘virtual reality’ where the goal is to feel immersed in an artificial environment - in telepresence the display shows the real world whereas in VR it shows a computer generated world. The practice of creating a simulated immersive environment for entertainment has been around since the early 19<sup>th</sup> Century in the form of cycloramas and the like, an interesting survey of these early simulations can be found in [34] However a leap forward in this concept appeared in Ray Bradbury’s short story “The Veldt” which was published in the Saturday Evening Post in 1950 and subsequently included in his collection “The Illustrated Man” in 1951 [35]. In this story a nursery is described which anticipates the factual CAVE like environments of today although the nursery in Bradbury’s story is much more sophisticated as it creates an artificial environment directly from the thoughts of its users. The children in the story had imagined an African scene that worried their parents, as the parents enter the room Bradbury writes: “The walls were blank and two-dimensional. Now as George and Lydia Handley stood in the centre of the room, the walls began to purr and recede into crystalline distance, it seemed, and presently an African Veldt appeared, in three dimensions, on all sides, in colour, reproduced to the final pebble and bit of straw. The ceiling above them became a deep sky with a hot yellow sun”. And then: “Now the hidden odorophonics were beginning to blow a wind of odour at the two people in the middle of the baked veldtland...”

Then in 1952 in a short story titled “Bridge” the science-fiction author James Blish [36] describes a construction engineer carrying out work in the tempestuous and hostile atmosphere of Jupiter as though he is there, while in reality he is physically present on a satellite – “Jupiter V”. His vicarious presence on Jupiter is achieved through technological mediation that includes the use of a head mounted display. Today, over six decades later, we would call this worker’s experience telepresence. This is a short

passage from the beginning of the story when the worker is just completing a spell of driving a squat “beetle” vehicle along the ice-bridge that is being built in an environment of howling storms and crushing pressures.

“In the momentary glare, however, he saw something – an upward twisting of shadows, patterned but obviously unfinished, fluttering in silhouette against the hydrogen cataract’s lurid light. The end of the Bridge. Wrecked. Helmuth grunted involuntarily and backed the beetle away. The flare dimmed; the light poured down the sky and fell away into the raging sea below. The scanner clucked with satisfaction as the beetle recrossed the line into Zone 113. He turned the body of the vehicle 180<sup>0</sup>, presenting its back to the dying torrent. There was nothing further he could do at the moment on the Bridge. He scanned his control board – a ghost image of which was cast across the scene on the Bridge – for the blue button marked *Garage*, punched it savagely, and tore off his helmet. Obediently the Bridge vanished.” Then a few lines later: “The abrupt transition from the storm-ravaged deck of the Bridge to the quiet, placid air of the control shack on Jupiter V was always a shock. He had never been able to anticipate it, let alone become accustomed to it; it was worse each time, not better. He put the helmet down carefully in front of him and got up, moving carefully upon shaky legs; feeling implicit in his own body the enormous pressures and weights his guiding intelligence had just quitted.”

Thus as well as a Jovian telepresence robot being suggested, we also have a head mounted display (HMD), augmented reality, and immersion withdrawal symptoms [37] all described almost a decade before the first physical HMD. I say this because I have found no evidence of the previously noted McCollum patent of 1943 actually being built. However a rapid succession of real world HMDs now appear. In 1957 Morton L Heilig filed a patent for a “Stereoscopic–Television Apparatus for Individual Use” and this was granted in Oct 4<sup>th</sup> 1960 [38]. In the 10<sup>th</sup> of November 1961 edition of *Electronics* a report was presented on C P Comeau’s and J S Bryan’s Head-sight television system for the Philco Corporation [39]. And finally, for displaying computer generated rather than televised images, we have Ivan Sutherland’s 1968 paper in which he describes the development of “A head-mounted three dimensional display”. This included hardware and software development to create an immersive experience for the user although at this time miniature CRTs still had to be used [40].

Today the concept of being able to immerse ourselves in computer generated environments to the extent that these virtual world appear real has long been in science fiction and popular culture from previously noted *The Veldt* in 1951, to *Counterfeit World* in 1965 [41] to *Neuromancer* in 1984 [42] to *Snow Crash* in 1992 [43] to the *Matrix* film in 1999, and now is quite common in books, films, and television dramas. However the ability to immerse ourselves in the real world at a remote location has not been so popular even although the concept appeared in fiction before ‘virtual reality’. We have already noted the examples of telepresence from 1938 and 1952 and more recent examples would be in the graphic novel series “*The Surrogates*” [44] adapted as the 2009 film of the same name where the population in many cities live their lives through telepresence robots that are attractive perfected versions of their own bodies, similar to Wellman’s story of seven decades earlier.

In the early stories it is interesting to note the use of each author’s own terminology to describe their imagined technologies. In *The Robot* and the *Lady* we find “goggle like televisors”, in *Brave New world* “feelies”, in *The Veldt* “odorophonics”,

and in Bridge “ultraphone “eyes”. Therefore although Persky’s term *television* had been widely adopted half a century after it had been first introduced, other terminology useful to telepresence was still being developed. For example the term teleoperation was not used until 1966 by E.G. Johnsen [45], and the word “telepresence” was first observed in print when used by Marvin Minsky in an essay for OMNI magazine in June 1980 [46].

## 4 Conclusion – Transparent Telepresence

Today we have many elements developed but not yet what I term a ‘transparent telepresence’ system where the user experiences full presence in the remote environment [47, 48]. However there are two relevant contemporary manifestations of earlier mentioned science fiction stories. The first is the teleoperation from a space station of a robot on Earth, as in the previously noted Heinlein story *Waldo*. This was reported in November 2012 [49] when a small robot on Earth was driven from the International Space Station. However in this NASA European Space Agency collaboration a laptop was used in the space station rather than a master arm. Secondly, extrapolating from the aforementioned 1938 Wellman story to the 2009 motion picture ‘*Avatar*’ we can imagine being fully immersed and in control of a surrogate body. One of the most relevant current research projects that approaches this ideal is the European Integrated Project VERE (Virtual Embodiment and Robotic Re-Embodiment) [50]. Some aspects of this attempt not only aural and visual telepresence within, but also mind control of a remote robot. This multi-million euro project began in June 2010 and will run for 60 months until 2015.

## References

1. McClelland, S. (ed.) *Future Histories*. Pub. Nokia and Horizon House (June 1997)
2. Bell, A.G.: An Improvement in Telegraphy; US Pat. No. 174,465 (filed February 14 1876), (granted March 7, 1876)
3. Edison T.: Improvement in Phonograph or Speaking Machines; US Pat. No. 200,521 (filed December 24, 1877), (granted February 19, 1878)
4. *Scientific American*, pp. 384–385 (December 22, 1877)
5. Nipkow, P. G.: An electric telescope for the electric reproduction of illuminating objects. German Pat. No. 30105, (granted January 15, 1885) (retroactive January 6, 1884).
6. Verne, J., Verne, M.: In the Year 2889. first published in *The Forum* February 1888. Available in “*The Works of Jules Verne*” Published by Raleigh St. Claire Books (June 3, 2009)
7. Wells H.G.: *The Remarkable Case of Davidson’s Eyes*. Pall Mall Budget (March 28, 1895), available online from The Literature Network, <http://www.online-literature.com/wellshg/2867/>
8. Wells, H.G.: *The Crystal Egg*, *The New Review* (May 1897), web edition available online from University of Adelaide, [http://ebooks.adelaide.edu.au/w/wells/hg/crystal\\_egg/](http://ebooks.adelaide.edu.au/w/wells/hg/crystal_egg/)
9. Rosheim, M.E.: *Robot Evolution*. John Wiley and Sons Inc., New York (1994)

10. Perskyi, P.M.C., Television Au Moyen De L'electricite. In: 1st International Electricity Congress, Paris (August 25, 1900)
11. Bidwell, S.: Telegraphic Photography and Electric Vision. *Nature* 78(2014), 105–106 (1908)
12. Campbell-Swinton, A.A.: Distant Electric Vision. *Nature* 78(2016), 151 (1908)
13. Baird J.L., Day: A system of transmitting views portraits and scenes by telegraphy or wireless telegraphy. UK Pat. No.GB222604 (October 9, 1924)
14. Zworykin, V.K.: Television System, US Pat. No. 2,141,059 (December 20, 1938)
15. Farnsworth, P.:A Television System US Pat. No.1,773,980. (August 26, 1930)
16. Farnsworth P.: A Television Receiving System. US Pat. No. 1,773,980 (August 26, 1930)
17. Huxley, A.: *Brave New World.. First Edition 1932 Published by Chatto and Windus (London). Quote from pp. 145 – 147 Vintage Classics edition (2007)*
18. <http://www.stereoscopy.com/faq/wheatstone.html> (retrieved March 9, 2012)
19. Heilig, M.L.: El Cine Del Futuro: The Cinema of the Future, originally published in *Espacios* 23–24 (1955), reprinted with translation Presence, Vol.1(3), pp. 279–294. MIT Press (1992)
20. Wellman, M.W.: The Robot and the Lady. Published October 1938 in 'Thrilling Wonder Stories' Text in this paper extracted from 'technovelgy' web pages, <http://www.technovelgy.com/ct/content.asp?Bnum=1221> (March 15 2012)
21. Heinlein, R.: Waldo. originally published under the pseudonym Anson MacDonald in 'Astounding Magazine' (August 1942)
22. Gallun, R.Z.: A Menace in Miniature. *Astounding Stories* (October 1937)
23. Clarke, A.C.: V2 for Ionosphere Research?, in Letters to the Editor, *Wireless World* (February 1945)
24. [http://en.wikipedia.org/wiki/Herman\\_Potocnik](http://en.wikipedia.org/wiki/Herman_Potocnik) (accessed March 19, 2012)
25. Henkes, P.M.: Three Dimensional Engraving and Allied Pantograph Machine, US Pat. No. 2,161,709, (filed June 3, 1938), (granted April 30, 1940)
26. Pollard, W. L. V.: Position Controlling Apparatus US Pat No 2,286,571 (filed April 22,1938) (granted June 16,1942)
27. Roselund, H.A.: Means for Moving Spray Guns or Other Devices Through Predetermined Paths, US Pat No 2,344,108 (filed August 17, 1939) (granted March 14, 1944)
28. Hull, H.L., Hartman, W.C., Goertz, R.C.: Handle Control System, US Pat No 2,414,102 (filed July 23, 1941) (issued January 14, 1947)
29. Goertz, R.C., Thompson, W.M.: Electronically Controlled Manipulator. *Nucleonics*, 46–47 (1954)
30. Hull, H.L., Goertz, R.C.: Positional Control System, US Pat No 2,526,665 (filed January 24,1942) (granted October 24, 1950)
31. Peoples, J.R., Schelb, R., Goertz, R.C.: Servo System and Control Thereof, US Pat No 2,466,041 (filed March 30th 1943) (granted April 5, 1949)
32. Goertz, R.C.: Remote Control Manipulator, US Pat No 2,632,574 (filed December 16, 1949), (granted March 24, 1953)
33. De N McCollum, H. J., McCollum, T.: Stereoscopic Television Apparatus, US Pat No 2,388,170, (filed April 15, 1943), (granted October 30, 1945)
34. Judith, M.: Fly Me to the Moon: A Survey of American Historical and Contemporary Simulation Entertainments. *Presence* 6(5), 565–580 (1997)
35. Bradbury, R.: The Veldt, in 'The Illustrated Man', first published in UK 1952 by Rupert Hart-Davis Ltd., *Flamingo Modern Classic edition*, pp.15–32 (1995)

36. Blish, J.: *Bridge*, Astounding Science Fiction, pp. 57–82. Pub. Street and Smith publications, New York (1952)
37. Regan, C.: An Investigation into Nausea and Other Side-effects of Head-coupled Immersive Virtual Reality. *Virtual Reality* 1(1), 17–32 (1995)
38. Heilig, M.L.: Stereoscopic-Television Apparatus for Individual Use, US Pat No 2,955,156 (filed May 24, 1957) (granted October 4, 1960)
39. Comeau, C.P., Brian, J.S.: Headsight Television System Provides Remote Surveillance. *Electronics*, pp. 86–90 (November 10, 1961)
40. Sutherland, I.E.: A head-mounted three dimensional display. In: Fall Joint Computer Conference, AFIPS Conference Proceedings, vol. 33, pp. 754–764 (1968)
41. Galouye, D.F.: *Counterfeit World*, Pub. Victor Gollancz Ltd. (1965)
42. Gibson, W.: *Neuromancer*, Pub. Victor Gollancz Ltd. (1984)
43. Stephenson, N.: *Snow Crash*, Pub. Bantam Books (1992)
44. Venditti, R., Weldele, B.: *The Surrogates*, graphic novel series, Pub. Top Shelf Productions (2005-2006)
45. Johnsen, E.G.: Telesensors, teleoperators and telecontrols for remote operations. *IEEE Transactions on Nuclear Science* NS13, 14–21 (1965); (Originally presented at the 12th Annual Nuclear Science Symposium, San Francisco, California, October 18-20, 1965).
46. Minsky, M.: Telepresence. *OMNI Magazine* (June 1980)
47. Mair, G.: Transparent telepresence research. *Industrial Robot* 26(3), 209–215 (1999)
48. Mair, G.: Towards Transparent Telepresence. In: Shumaker, R. (ed.) *HCI 2007 and ICVR 2007*. LNCS, vol. 4563, pp. 300–309. Springer, Heidelberg (2007)
49. [http://www.nasa.gov/home/hqnews/2012/nov/HQ\\_12-391\\_DTN.html](http://www.nasa.gov/home/hqnews/2012/nov/HQ_12-391_DTN.html) (retrieved February 27, 2013)
50. <http://www.vereproject.eu> (retrieved February 25, 2013)

# High Presence Communication between the Earth and International Space Station

Tetsuro Ogi, Yoshisuke Tateyama, and Yosuke Kubota

Graduate School of System Design and Management, Keio University  
4-1-1 Hiyoshi, Kohoku-ku, Yokohama 233-8526, Japan  
ogi@sdm.keio.ac.jp,  
tateyama@sdm.keio.ac.jp, yosuke-k21@z2.keio.jp

**Abstract.** In this study, in order to realize high presence communication with the astronaut who is staying on the ISS, the experiment on remote communication using the technologies of 2D/3D conversion, immersive dome display, and sharing space among multiple sites were conducted. In this case, biological information such as electrocardiogram, thermal image, and eye movement were measured to evaluate the sense of presence, and the tendency that the user felt the high presence sensation when experiencing the high resolution three-dimensional stereo image. From these results, we can understand that high presence communication between the earth and the ISS was realized.

**Keywords:** Tele-immersion, High Presence Sensation, Biological Information, 2D/3D Conversion, International Space Station.

## 1 Introduction

Recently, outer space has become not only the space that is flown through by the rocket or space shuttle but also habitation space where human can stay for a long time, according to the development of ISS (International Space Station) [1]. Therefore, it is one of the important demands to realize high presence communication between the human who is on the earth and the astronaut who is staying on the space station. In order to solve such a demand, the researches about the high presence communication using the virtual reality or tele-immersion technologies are expected to be an effective approach.

In November, 2012, the authors got an opportunity to perform a communication event with the ISS, and we conducted an experiment on high presence communication between the earth and the ISS in this event. In the communication event, seven sites were connected to the ISS simultaneously, and the graduate students and the elementary and junior high school students asked questions from Keio University to astronaut Akihiko Hoshide staying on the ISS. In this event, in order to realize high presence communication, we applied several technologies such as high resolution 3D image, immersive dome display, and sharing space among multiple sites. This paper describes the experiment about the high presence communication between the earth and the ISS that was conducted in this event.



## 2 High Presence Communication

In the communication event with the ISS, the image sent from the ISS was received at several sites simultaneously and it was shared among all sites while performing each event respectively.

As a conversation place where the students talk to the astronaut on the ISS, CDF (Concurrent Design Facility) room in Keio University was used [2], where the symposium entitled "Information Technology and Space Age" was performed. In addition, extension lecture about space engineering was held in the main hall in the same building as the CDF room, and the events for children were held at citizen halls in Setagaya-ku and Minamisanriku-cho. The video image transmitted from the ISS was once received in the CDF room, and it was distributed to other sites. Moreover, the air dome display that was built temporarily in Keio University, and the conference rooms at the University of Tokyo and Kyoto University were also connected to the CDF room through the network, and the video image edited in the CDF room was distributed to each site.

In this study, the experiment was conducted for the purpose of realizing high presence communication between the earth and the ISS. Although various systems that use stereo image, high resolution image, or large screen image have been developed to represent or transmit high presence sensation, the equipment that can be used in the current ISS is restricted. Therefore, in this experiment, we gave up realizing high presence mutual communication and aimed at generating high presence sensation only at the earth side.

As the concrete methods about the high presence communication, the following technologies were applied;

1. 2D/3D conversion
2. Immersive dome display
3. Sharing space among multiple sites.

## 3 Network System

Figure 1 shows the construction of the network system that was built in this experiment. Though the network for the image transmission from the ISS to NASA and JAXA (Japan Aerospace Exploration Agency) is always secured and monitored, the network to other places must be constructed by ourselves. In this experiment, the network environment in which the communication image that is transmitted from the ISS to JAXA Tsukuba Space Center can be sent to Keio University through SINET (Science Information Network) or Tsukuba WAN was constructed. Since the communication load in SINET changes depending on the period of time, another path of the network using Tsukuba WAN and JGN-X (Japan Gigabit Network eXtreme) was prepared for backup. These networks were connected to the CDF room in Keio University through WIDE (Widely Integrated Distributed Environment) network.

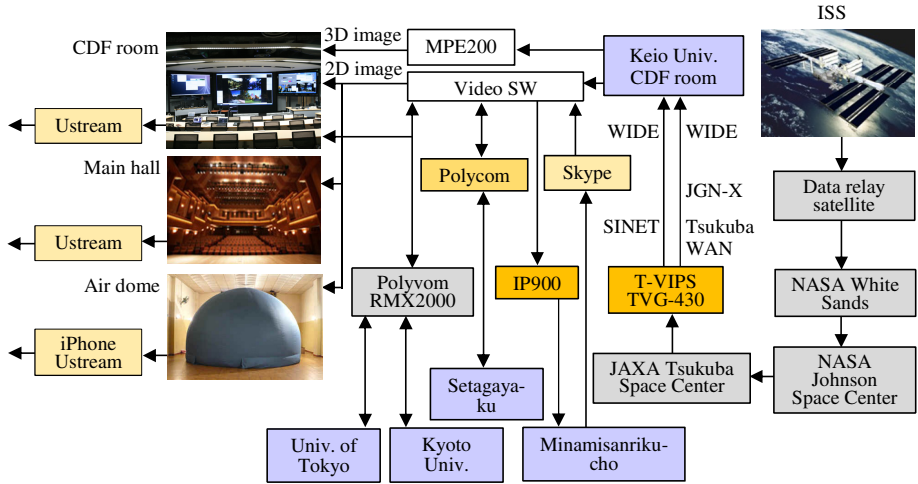


Fig. 1. Network construction of the remote communication between the earth and ISS

As for the method of the image transmission, IP transmission equipment, video conferencing system, IP TV phone, and video streaming service were used for different purposes. Since the image that was transmitted from JAXA Tsukuba Space Center to the CDF room at Keio University was used for the conversation between the earth and the ISS directly as well as for the shared image among all sites, T-VIPS TVG-430 of the IP transmission equipment based on JPEG2000 format was used in consideration of the image quality and real-time processing.

The video image received at the CDF room was then distributed to other sites. Since the main hall and the air dome were placed in the same building as the CDF room, the video image was transmitted from the CDF to two sites using the coaxial cable via a distributor. In addition, the video images were distributed mutually among three sites of the CDF room, the University of Tokyo, and Kyoto University using a multi-point HD video conference connection server Polycom RMX2000. At the citizen hall in Setagaya-ku, HD video conferencing system Polycom HDX8005 was used independently, in order to exchange the images mutually between the CDF room and the citizen hall and to use the image of the other site in each event. Though the image of the ISS was transmitted from the CDF room to the citizen hall in Minamisanriku-cho using the H.264 format by the IP transmission equipment Fujitsu IP900, the image captured at Minamisanriku-cho was sent to the CDF room using Skype of IP TV phone to transmit only the atmosphere of the hall.

Moreover, the video images of the events held in the CDF room and main hall at Keio University were broadcasted on the Internet using Ustream video streaming service. The video image captured at the air dome in Keio University was broadcasted using the iPhone Ustream, and this image was also used for the communication between the dome display and the other sites. Namely, in this system, the images were transmitted using the appropriate methods in consideration of the network bandwidth, time delay, the usability of the communication equipment, and so on.

## 4 Communication Using 3D Image

First, in this study, in order to realize high presence communication, high resolution 3D image was used in the CDF room. Since a stereo camera cannot be installed in the ISS, 2D/3D conversion technology was applied to the transmitted monocular HD image in real time and the three-dimensional stereo image was generated. The communication time given by JAXA was about 20 minutes, and data relay satellite was switched in the middle of the communication. Therefore, the transmitted two-dimensional video image was displayed without conversion for 10 minutes of the first half, and the converted three-dimensional video image was displayed for 10 minutes of the second half. The image processor SONY MPE-200 was used for 2D/3D conversion, and the generated three-dimensional stereo image was projected onto the 180-inch screen using the stacked 4K projectors of SONY SRX-S110 through the polarizing filter.



©JAXA

**Fig. 2.** Conversation between student and astronaut in ISS using 3D image

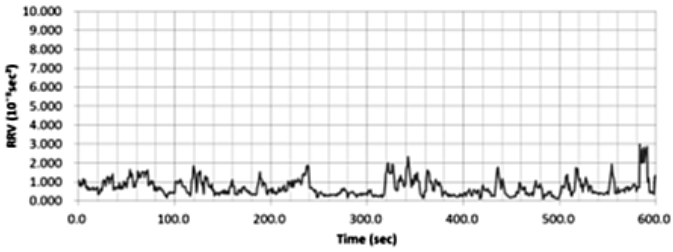
In general, in the method of 2D/3D conversion, three-dimensional scene is constructed by recognizing objects in the monocular image and adding the information of binocular parallax to each object. In this case, the added information of the binocular parallax is not necessarily accurate, because it is generated automatically according to the conversion algorithms such as the focal point analysis. In this experiment, since the full-scale model of Japanese experiment module "Kibo" in the ISS exists in JAXA Tsukuba Space Center, it was possible to measure the size of the model and to adjust the value of the binocular parallax so that the image can be represented as real scale. In this case, the parameters for binocular parallax were adjusted using the video image that was recorded in the previous event, so that the image of the inner space of "Kibo" was represented as three-dimensional real scale

image. However, the image of astronaut Hoshide that was generated in the real-time communication was represented as larger size, because the positions of the astronaut and the video camera in the actual communication were unknown. Figure 2 shows the three-dimensional stereo image of astronaut Hoshide that was projected onto the 180-inch screen. In the CDF room, the graduate students and the elementary and junior high school students talked with astronaut Hoshide while looking at the three-dimensional stereo image.

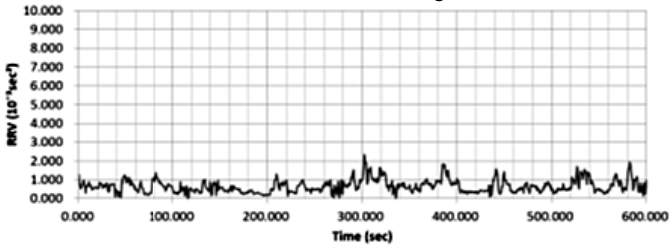
The purpose of this experiment includes the evaluation of the high presence sensation that was represented in the remote communication as well as the realization of the real-time communication between the earth and the ISS. However, the questions of “What is the high presence sensation?” and “How the high presence sensation can be measured?” are not clearly defined. In order to solve this problem, we have examined the method of evaluating the high presence sensation based on the measurement of biological information. From the past research, we have found that when the persons are experiencing high presence images, the value of RRV (variance of R-R intervals) measured by electrocardiograph becomes lower [3], the nose temperature measured by thermal camera does not fall [4], and the view point measured by eye tracker moves frequently. This means that when the person is experiencing the high presence sensation, he is concentrating his attention, feeling few mental stresses, and interested in a lot of objects.

In this experiment, we measured the electrocardiogram, the thermal image, and the eye movement of the subject who was sitting on the center seat of the front row in order to evaluate the presence felt by the subject, and compared the sensation when he was seeing the two-dimensional image and three-dimensional image of the astronaut.

Figure 3 shows the change in the value of RRV. Though the average of RRV was  $0.715 (10^{-3}\text{sec}^2)$  when seeing the two-dimensional image, it was  $0.657 (10^{-3}\text{sec}^2)$  when seeing the three-dimensional image, and there was significant difference between them. Figure 4 shows the change in the nose temperature and forehead temperature. The average of the difference between nose temperature and forehead temperature was 0.276 (degrees) and 0.141 (degrees) when seeing the two-dimensional image and three-dimensional image, respectively. And there was significant difference between them. As for the eye movement, the frequency of view point movement was 1.064 (times per second) and 1.147 (times per second) when seeing the two-dimensional image and three-dimensional image, respectively. And there was a tendency that the view point moves a little frequently when seeing the three-dimensional image. From these results, we can recognize that when the subject was experiencing the three-dimensional image of astronaut Hoshide, the value of RRV became low, the nose temperature did not fall, and view point moved frequently. Namely we can understand that the subject experienced the high presence communication with the astronaut Hoshide by using the high resolution three-dimensional stereo image.

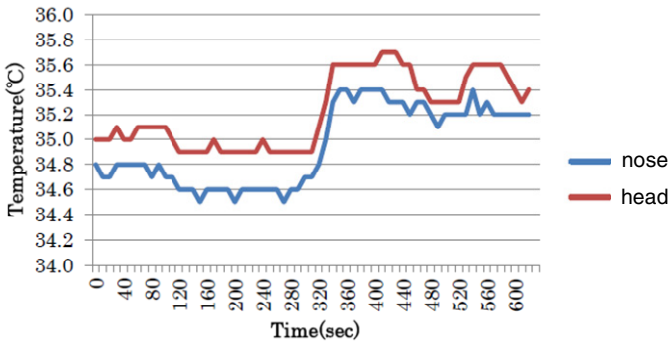


(a) 2D image

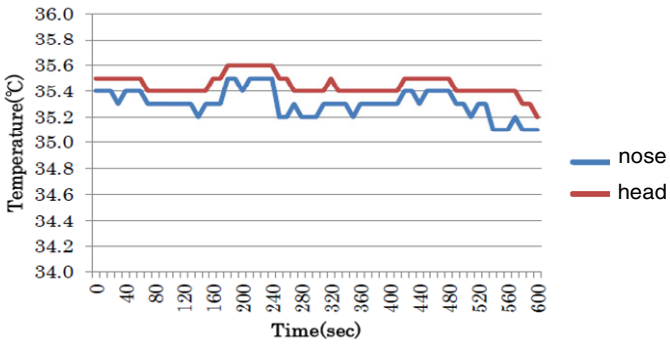


(b) 3D image

**Fig. 3.** RRV value when seeing (a) 2D image and (b) 3D image



(a) 2D image



(b) 3D image

**Fig. 4.** Nose and head temperature when seeing (a) 2D image and (b) 3D image

Moreover, as subjective evaluation, we conducted a questionnaire survey to the participants in the communication event performed in the CDF room. In the questionnaire, the questions about the communication using the two-dimensional image and the three-dimensional image shown in Table 1 were asked. Participants were asked to answer each question using five-grade system such as "agree (+2), somewhat agree (+1), neither agree nor disagree (0), somewhat disagree (-1), disagree (-2)". In the table, the average values and standard deviations of 38 persons who answered the questionnaire and the results of t-test for each question are shown. From the result, we can see that there is significant difference in the questions of "I felt that astronaut was talking to me" and "I felt that I was in the same space with astronaut". Therefore, we can understand that the high presence sensation was represented by using the high resolution three-dimensional image from the subjective evaluation.

**Table 1.** Result of questionnaire about 2D and 3D image

<i>question</i>	<i>3D image</i>	<i>2D image</i>	<i>t-test</i>
I felt three-dimensional sensation.	1.16	-0.34	P=0.000
I concentrated attention.	1.21	0.71	P=0.026
I felt excited.	0.89	0.21	P=0.047
I felt tired.	-0.34	-0.92	P=0.029
I felt that astronaut was talking to me.	0.63	-0.08	P=0.007
I felt that I was in the same space with astronaut.	0.58	-0.13	P=0.006

(-2) disagree -- (0) -- agree (+2)

## 5 Immersive Dome Display

Next, the immersive dome display was used to represent the high presence sensation. In the dome environment, it is known that the user can experience immersive sensation from the frameless image with wide field of view. In addition, it is also known that the user can feel three-dimensional sensation from the monocular image with wide field of view without wearing 3D glasses [5]. In particular, in the content of the remote communication with the ISS, we can expect the effect that the planetarium-like environment of the dome display reinforces the sense of presence by making the user imagine the space.

In this experiment, the air dome system of CUBEX Dome with 5 meters in diameter was built temporarily in the same building as the CDF room. In this system, the image was projected onto the dome screen by using one projector to which 180 degrees fish-eye lens was attached. Though, in general, the image projected by the fish-eye projector contains large distortion, in this system, the transmitted image was directly projected onto the screen without correcting the distortion, because the distortion based on the fish-eye lens was reduced by the shape of the dome screen.

Figure 5 shows the appearance of the air dome system that was installed in this experiment and figure 6 shows the image of astronaut Hoshide that was projected onto the dome screen. Though the number of users who can experience the air dome simultaneously was about ten, the comments such as "The immersion felt in the dome display was suitable for the representation of the space." and "The size of the

dome display was suitable to represent the image of the astronaut with high presence sensation." were gotten from the users. From these comments, we can see that the user experienced the high presence image effectively in the dome environment. Although we planned to use the large-scale planetarium at the beginning, it is thought that this air dome was a suitable size to represent the internal scene of the ISS.



**Fig. 5.** Air dome with 5 meters in diameter built in the building



©JAXA

**Fig. 6.** Image of astronaut Hoshide projected onto the dome screen

## 6 Sharing Space among Multiple Sites

This communication event was performed by connecting seven sites, such as the CDF room, the main hall, and the air dome in Keio University, the citizen halls in Setagaya-ku and Minamisanriku-cho, and the conference rooms in the University of Tokyo and Kyoto University through the network. Though the several events of the symposium, the extension lecture, and the quiz were performed in the CDF room,

main hall at Keio University, and the citizen halls in Setagaya-ku and Minamisanriku-cho respectively, whole event was also constructed in all sites by sharing information based on the image transmission. In particular, the sensation of sharing space was generated by displaying the same image at all sites while performing the communication between the earth and the ISS.

We can say that the participants experienced a kind of parallel reality that consisted of several sites including the ISS. Parallel reality is a concept in which the user can experience several real worlds simultaneously by using various kinds of image transmission technologies. In this experiment, space sharing among multiple sites was realized by using various image transmission technologies such as IP transmission equipment, video conferencing system, IP TV phone, and video streaming service.

Particularly, since the CDF room that was used for the conversation with the ISS consists of 180-inch 4K3D screen and two 108-inch LCD monitors placed at both sides, three-dimensional stereo image and other two kinds of images can be displayed at the same time. Moreover, since the 4K screen can display four divided high definition images, it can be used as multi-display environment where a total of six kinds of images can be displayed. Figure 7 shows that the users were experiencing the events of multiple sites simultaneously by looking at the images that were sent from other sites.

In this experiment, the participants in each hall were able to experience the events in other halls simultaneously by connecting each site mutually based on the image transmission, and the sensation of sharing space among multiple sites was generated. Namely, we can understand that this experimental event generated the consciousness of being experiencing one of the real worlds that includes the space environment for many participants.



Fig. 7. Experience of sharing space using multi-display environment

## 7 Conclusions

In this study, the experiment on high presence communication between the earth and the ISS was conducted. In the experiment, the technologies of 2D/3D conversion, immersive dome display, and sharing space among multiple sites were applied to



realize the high presence communication. In particular, in the experiment of 2D/3D conversion, the sensation of the presence felt by the user was verified based on the measurement of the biological information.

Although each technology of the image transmission used in this experiment was completed technology, it was not easy to use them simultaneously and manage the whole system, and some small troubles such as the interruption of the sound happened in the communication event. Future work will include acquiring more experiences of the communication experiment and establishing the technology that can be used more easily even in the space station.

**Acknowledgements.** This research was supported by Keio Gijuku Academic Development Funds and G-COE (Center of Education and Research of Symbiotic, Safe and Secure System Design) program at Keio University. We thank Prof. Naohiko Kohtake, Prof. Naohisa Ohta, and Prof. Akira Kato of Keio University, Mr. Yutaka Kaneko of JAXA, Mr. Tsuyoshi Doi of Yomiuri Shinbun, and other event staff members for their support.

## References

1. Okami, Y.: Space Station and Supporting Engineering. *Journal of the Institute of Electrical Engineers of Japan* 121(3), 199–202 (2001)
2. Ogi, T., Tsubouchi, D.: Development of Concurrent Design Environment Using Super High Definition Image. In: 2010 Asian Conference on Design & Digital Engineering (ACDDE 2010), pp. 85–88 (2010)
3. Hirose, M., Nakagaki, Y., Ishii, T.: Application for Operational Management using R-R Intervals of ECG Measured by Optical Sensor. *Proc. of the Japan Society of Mechanical Engineers* 884(5), 82–84 (1984)
4. Toma, T., Ogi, T.: Influence of Live Classic Concert on Stress and Relaxation in 3D High Presence Environment. In: Asian Conference on Design and Digital Engineering (ACDDE 2011 Proceeding II), pp. 503–506 (2011)
5. Ogi, T., Tateyama, Y., Lee, H., Furuyama, D., Seno, T., Kayahara, T.: Creation of Three Dimensional Dome Contents Using Layered Images. In: The 1st International Symposium on Virtual Reality Innovations (IEEE ISVRI), pp. 253–258 (2011)

# Effects of Visual Fidelity on Biometric Cue Detection in Virtual Combat Profiling Training

Julie Salcedo<sup>1</sup>, Crystal Maraj<sup>1</sup>, Stephanie Lackey<sup>1</sup>, Eric Ortiz<sup>1</sup>,  
Irwin Hudson<sup>2</sup>, and Joy Martinez<sup>1</sup>

<sup>1</sup> University of Central Florida, Institute for Simulation and Training, Orlando, FL  
{jsalcedo, cmaraj, slackey, eortiz, jmartine}@ist.ucf.edu

<sup>2</sup> U.S. Army Research Laboratory – Human Research Engineering Directorate  
Simulation and Training Technology Center, Orlando, FL  
irwin.hudson@us.army.mil

**Abstract.** Combat Profiling involves observation of humans and the environment to identify behavioral anomalies signifying the presence of a potential threat. Desires to expand accessibility to Combat Profiling training motivate the training community to investigate Virtual Environments (VEs). VE design recommendations will benefit efforts to translate Combat Profiling training methods to virtual platforms. Visual aspects of virtual environments may significantly impact observational and perceptual training objectives. This experiment compared the effects of high and low fidelity virtual characters for biometric cue detection training on participant performance and perceptions. Results suggest that high fidelity virtual characters promote positive training perceptions and self-efficacy, but do not significantly impact overall performance.

**Keywords:** Biometric Cue Detection, Visual Fidelity, Virtual Training.

## 1 Introduction

Combat Profiling training provides Warfighters with a valuable observational and decision-making skill set for dominating within the ever-changing irregular warfare environment. The U.S. Marines' Combat Hunter program provides Combat Profiling instruction, which trains Warfighters to observe the human terrain and establish a baseline for behavior along six domains including: atmospheric, biometrics, geographics, heuristics, kinesics, and proxemics [1]. Behavioral anomalies are analyzed to determine if a potential threat requiring further action is present. Combat Profiling skills equip Warfighters to be preemptive, rather than reactive, in their detection and mitigation of threats [2]. The apparent value of Combat Profiling has increased desires to expand the accessibility of training. Current research and development efforts are making strides toward cost-effective Combat Profiling training solutions including virtual simulation.

Design factors related to visual aspects in the Virtual Environment (VE) (e.g., graphics, visual fidelity, textures, 3D modeling, etc.) may significantly impact visual

design requirements related to observational and perceptual training objectives in the Combat Profiling domain. This research aimed to identify an acceptable level of visual fidelity to elicit positive trainee perceptions and increased performance with virtual Combat Profiling applications.

Visual fidelity refers to the visual similarity of a representation to the original and the ability of that representation to communicate the intended concept [3]. In graphics-based computer applications, such as interactive VEs, users may quickly compare a current image to one previously presented, but will rarely attend to direct comparison of visuals [3]. This implies that users distinguish a visual element and interpret its significance for the current state of the system (e.g., application, simulation, or scenario) based on its similarity to the target representation and intended meaning. This also suggests that user perception of visuals play a role in the discrimination and interpretation of these images.

Combat Profiling skills are predominantly observational and perceptual. For example, trainees receive instruction on how to identify potential threats by observing various cues that may indicate a target's intent or emotional state. Therefore, in a virtual Combat Profiling training setting, inability to portray the intended cue and meaning may result in negative training and poor performance. When depicting domain relevant human emotions in virtual characters, evoking accurate user perceptions of the intended emotion is a critical design implication. Users perceive emotions differently depending on the character features available to represent emotions [4]. Communicative cues used in human-to-human contact appear to translate into interactions between humans and virtual characters [5]. Therefore, users may better interpret intended emotions in characters designed with more human-like features and expression capabilities [4], [6]. Some facial features and behaviors that affect perceptions of emotion include: wrinkling, blushing, sweating, and tears [4-7].

Each domain of Combat Profiling likely varies in visual requirements for virtual training; therefore, research is needed to investigate the dynamics of each domain separately. This research initiative addresses the biometrics domain. The biometrics domain involves involuntary behaviors exhibited by the body during various emotional states. Examples include: sweating, flushing, blushing, pallor, visible veins, and heart rate [8]. The scope of this research focuses on two biometric cues addressed in current Combat Profiling training—facial sweat and facial flushing. In Combat Profiling training, trainees are taught the significance of these cues and how to distinguish them when attempting to identify potential threats among the human terrain.

Sweat is the moisture secreted by the sweat glands of the skin. There are two types of sweat: thermal and emotional [7]. Within Combat Profiling, the focus is on emotional sweating. Emotional sweating is caused by nervousness and anxiety or an intense emotional state, such as anger or rage [7-11]. Threatening individuals may exhibit signs of emotional sweat if they are nervous about being caught, have anxiety about completing their attack, or are experiencing anger, hatred, or rage toward an adversary [8], [11]. Emotional sweating is also evident in non-threatening individuals that may be in distress. Distressed individuals may exhibit signs of emotional sweat if they are nervous or anxious around other individuals who pose a potential danger [11]. Visible sweat caused by emotion often appears as a collection of droplets or as a sheen on the forehead, upper lip, nose, and/or cheeks [5], [7].

Flushing is the reddening of the face, neck, chest, and ears and may appear when individuals feel shyness, shame, anger, embarrassment, rage, anxiety, distress, or are performing deceptive acts [10-11]. For Combat Profiling, the focus is on flushing caused by anger, rage, deceptive behavior, anxiety, or distress [8]. Threatening individuals may exhibit signs of flushing if they are attempting to commit deceptive acts or experiencing anger or rage. Non-threatening individuals may exhibit signs of flushing if they are anxious or distressed in the presence of others who may pose a danger [11]. Flushing caused by anger, rage, deceptive behavior, anxiety, and distress often appears as redness on the forehead, nose, cheeks, ears, neck, and/or chest [10].

The purpose of this research was to investigate participant perceptions of biometric cues in a VE. Results will have implications on the design of visual elements for modeling biometric cues in virtual Combat Profiling training. Specifically, this experiment targets visual fidelity of virtual characters used to model sweating and flushing cues. Visual fidelity was varied by manipulating the polygon count of the virtual models, a method consistent with previous visual fidelity research [3]. A high polygon count was considered the high fidelity condition and a low polygon count distinguished the low fidelity condition. The flushing and sweat cues were modeled on the virtual characters faces using shaders and texture mapping tools in Autodesk's 3ds\_Max software.

This experiment compared participants' perceptions of the realism and effectiveness of high versus low fidelity virtual characters for biometric cue detection training. Performance measures were collected to compare participants' abilities to identify the biometric cues in the high versus low fidelity virtual characters. It is hypothesized that there is a difference in participant's perception of high fidelity virtual characters than low fidelity characters. Additionally, there is a difference in performance in scenarios using high fidelity virtual characters than low fidelity characters. As experience and previous research indicate, the modeling of realistic facial flushing cues is easier to accomplish than modeling realistic sweat cues [6]. Therefore, it is also hypothesized that participants will have higher performance measures in scenarios containing only targets exhibiting the flushing cue versus scenarios with only sweat cues. Additionally, performance self-assessment and emotional intelligence measures were collected to determine if these individual differences were predictors of scenario performance.

## **2 Method**

### **2.1 Participants**

Sixty-six undergraduate students from the University of Central Florida were recruited for participation. Participants were U.S. Citizens and at least 18 years of age. Participants received class credit upon completion of the experiment. The sample population was selected as their skills and abilities reflect those of novice soldiers and can be used as a baseline for data collection [12].

### **2.2 Measures**

Measurement instruments for this experiment included a series of subjective questionnaires, and performance logging during experimental scenarios. The

demographics questionnaire included questions pertaining to general biographical information (i.e., age, gender), military experience, video-game experience, and computer competence. The Trait Emotional Intelligence Questionnaire-Short Form (TEIQue-SF) consisted of 30 self-report items measuring emotional intelligence across emotional intelligence subscales including: well-being, self-control, emotionality, and sociability [13]. The Performance Self-Assessment Questionnaire (PSAQ) is designed for the participant to judge their performance based on nine task-related questions. This measure was used as an indicator of self-efficacy in the analysis. The Training Perception Questionnaire (TPQ) gauged the participant's awareness, understanding, and perception of the virtual characters displayed, using a ten question multiple choice format.

### 2.3 Discrimination Task

The discrimination task involved identifying virtual characters that display the specified biometric cue (i.e., sweat or flushing) within the simulation platform. The platform utilized was the Mixed Initiative eXperimental (MIX) testbed [14]. The VE gave a first-person perspective of a patrol route in a desert village (Figure 1). The display of sweating, flushing or neutral cues on characters was randomized within three separate scenarios including: sweat only, flushing only, or a combination. The participant was tasked with identifying which characters displayed the sweat or flushing cue in either the high or low fidelity condition. Accuracy was determined as the number of biometric cues detected correctly out of the total cues present within the scenario.



**Fig. 1.** Participant view within the MIX testbed

For each scenario, there were two visual fidelity conditions: low and high. The virtual characters used in this experiment were procured from a third party vendor. The initial state of each model was considered the high fidelity version. These original models were imported into Autodesk's 3dsMax, a commercial-off-the-shelf

3D graphics software application, and modified to create the low fidelity versions using the optimization filter. The optimization filter is a feature used to modify the polygon count, or number of polygonal faces, of a model. Designers adjust the polygon count of characters in VEs in order to optimize processing speeds. However, this may be detrimental to visual fidelity as users may misidentify cues and features due to changes in the polygon count.



**Fig. 2.** High and low fidelity models of virtual flushing and sweat cues

For experimental purposes, the optimization settings for the original, high-polygon models (i.e., high fidelity) were set to ten and edited to five for the low-polygon versions (i.e., low fidelity). Figure 2 provides an example of one civilian model used to exhibit the sweat and flushing cues in the high and low fidelity conditions. Scenarios included models depicting male and female civilians, foreign military, and U.S. military. The number of polygonal faces varied depending on the details of each model including: skin texture, size of facial features, and the presence of adornments (e.g., hats, helmets, and scarves). Table 1 lists the range of polygonal faces of each character type.

**Table 1.** Virtual character polygon ranges

Character Type	Low Fidelity	High Fidelity
Female Civilian	2300-2500 polygons	3200-3800 polygons
Male Civilian	2100-3200 polygons	3400-4300 polygons
Foreign Military	5300-6000 polygons	7100-7400 polygons
U.S. Military	5400-8700 polygons	7900-11,000 polygons

## 2.4 Procedure

Upon arrival, the participant was asked to read the informed consent, which also served as a debriefing form at the end of the session. The participant was randomly assigned to a high or low fidelity condition. After reading the consent form, the instructions were to complete the demographics questionnaire and the TEIQue-SF. After questionnaire completion, the participant read a task training slide presentation.

Following the task training presentation, the participant completed three, randomized scenarios. For consistency, each participant was positioned 30 inches from the monitor, which was established as a comfortable viewing distance during pilot experimentation. Scenario duration was approximately 20 minutes each. After each scenario, the participant completed a PSAQ and TPQ. The participant was given a five minute break between the first and second scenario. At the end of the final scenario and questionnaires, the participant was debriefed and dismissed.

### 3 Results

An independent samples t-test was conducted to compare the TPQ ratings for high and low fidelity conditions (Table 2). There was a significant difference in TPQ scores for high fidelity and low fidelity conditions in the flushing scenario and combination scenarios with the high fidelity condition yielding higher ratings. Although the difference in TPQ scores was not significant for the sweat scenario ( $p=.056$ ), the descriptive statistics were consistent with results from the flushing and combination scenarios. Overall, the results suggest that individual's perception in the high fidelity condition were consistent with viewing high polygon models. Individuals in the low fidelity condition rated realism less certain and accurate.

**Table 2.** Training Perception Questionnaire (TPQ) independent samples t-test results

	High Fidelity		Low Fidelity		t	p	95% Confidence Interval	
	M	SD	M	SD			Lower	Upper
Flushing	2.94	.39	2.67	.48	(55)=2.48	.016	.056	.482
Sweat	2.90	.41	2.69	.48	(58)=1.95	.056	-.002	.434
Combo	2.93	.38	2.67	.40	(61)=2.68	.010	.067	.454

There was no significant difference between high and low fidelity conditions in PSAQ ratings for all three scenarios. Additionally, the performance scores (i.e., accuracy) revealed no significant difference between the high and low fidelity conditions. The percentage of correctly identified cues when compared to the total number of targets presented showed no significant difference between high and low conditions for all scenarios.

A multiple linear regression analysis was conducted to test if the PSAQ scores significantly predict the percentage of correctly identified cues. The results of the regression indicated that the PSAQ of the flushing scenario explained 15% of the variance in the flushing scenario;  $R^2=.17$ ,  $F(1, 63) = 10.57$ ,  $p=.002$ , and 18% of the variance in the sweat scenario;  $R^2=.21$ ,  $F(1, 63) = 15.95$ ,  $p<.001$ . However, the PSAQ was not a significant predictor of performance in the combination scenario.

A Spearman's rho correlation was computed to assess the relationship strength and direction between TPQ and PSAQ ratings for the high and low fidelity conditions collectively. There was a positive correlation between the two variables for the flushing scenario;  $r_s(66) = .390$ ,  $p=.001$ , the sweat scenario;  $r_s(66) = .513$ ,  $p<.001$ , and the combination scenario;  $r_s(66) = .417$ ,  $p=.001$ . There was also a positive correlation

between the two variables for the total ratings across all three scenarios;  $r_s(66) = .522$ ,  $p < .001$ . Overall, there was a moderate, positive correlation between TPQ ratings and PSAQ scores. Increases in ratings on the TPQ were correlated to increases in the scores on the PSAQ.

A multiple regression was used to test if personality traits predict the percentage of correctly identified cues by identifying the correct targets located in the flushing scenario. Results suggest that personality traits account for 12% of the variance;  $R^2 = .20$ ,  $F(6, 59) = 2.52$ ,  $p = .031$ . This indicates that personality traits may predict ability to correctly identify cues from the total number of correct flushing targets found. Notably, participants high in well-being significantly predicted the percentage of correctly identified and located cues within the flushing scenario;  $R^2 = .16$ ,  $F(1, 64) = 12.21$ ,  $p = .001$ .

An analysis of data also supported personality as a predictor of the participants' percentage of correctly identified cues to the total number of sweat targets. The regression results suggest that 17% of the variance is explained by personality;  $R^2 = .25$ ,  $F(6, 58) = 3.25$ ,  $p = .008$ . It was also found that personality is a predictor for correctly identifying sweat cues for the total number of sweat targets.

## 4 Discussion

Results of the TPQ indicate that participants had a more positive perception of training effectiveness and realism for the high fidelity virtual character condition over the low fidelity condition. It was also observed that a positive perception of the training tool correlated to higher self-assessments of performance, as indicated by the correlations between TPQ and PSAQ scores. Combining these observations suggests that higher visual fidelity in virtual biometric cue detection training may be associated with greater self-efficacy for the task. Previous research suggests that greater self-efficacy contributes to increased motivation, engagement, and training transfer [15-16]. Therefore, higher levels of visual fidelity for the biometric domain may contribute to increased motivation, engagement, and training transfer for virtual Combat Profiling training. Future research may investigate this further by comparing results from self-efficacy scales between high and low visual fidelity conditions.

Self-assessment ratings on the PSAQ were found to be significant predictors of performance for identifying flushing and sweat cues in a VE. This suggests that self-efficacy has a bearing on virtual biometric cue detection ability. Likewise, having TEI-Que scores high in psychological well-being was also a predictor of performance, which suggests that fostering a greater sense of well-being during virtual biometric cue detection training may result in a positive performance outcome. These findings are consistent with previous research indicating that self-efficacy and psychological well-being affect learners' behaviors and attitudes, which may contribute to higher or lower quality in learner performance [17]. Future research may directly assess the impact of virtual Combat Profiling training quality on Soldiers' self-efficacy and psychological well-being.

Although results suggest that there were no significant performance differences between high and low fidelity conditions, the emergence of self-efficacy as a potential motivation and engagement factor support the use of high fidelity models for the



biometric domain of Combat Profiling. Achieving a higher level of detail is no longer cost-prohibitive with the current technological advancements such as higher computer processing speeds and faster refresh rates of current video cards.

## 5 Conclusion

In anticipation of demands by the U.S. military to increase access to virtual Combat Profiling training and improve virtual training quality, this research sought to identify the impact of visual realism on trainees' abilities to detect biometric cues in a VE. The findings suggest that in the biometrics domain of Combat Profiling, it is recommended but not essential to include high fidelity virtual characters for the depiction of cues. It is further recommended that design requirements incorporate the use of high fidelity characters to improve the quality and perceptions of virtual training.

**Acknowledgements.** This research was sponsored by the U.S. Army Research Laboratory – Human Research Engineering Directorate Simulation and Training Technology Center (ARL HRED STTC), in collaboration with the Institute for Simulation and Training at the University of Central Florida. This work is supported in part by ARL HRED STTC contract W91CRB08D0015. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL HRED STTC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

1. Gideons, C., Padilla, F., Lethin, C.: Combat Hunter. *Marine Corps Gazette* 92(9), 79–84 (2008)
2. Schatz, S., Reitz, E., Nicholson, D., Fautua, D.: Expanding Combat Hunter: The Science and Metrics of Border Hunter. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (2010)
3. Watson, B., Friedman, A., McGaffey, A.: Measuring and Predicting Visual Fidelity. In: *28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 213–220. ACM, New York (2001)
4. Beer, J., Fisk, A., Rogers, W.: Recognizing Emotion in Virtual Agent, Synthetic Human, and Human Facial Expressions. In: *54th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 2388–2392. Sage, Santa Monica (2010)
5. De Melo, C.M., Kenny, P., Gratch, J.: Influence of Autonomic Signals on Perception of Emotions in Embodied Agents. *Applied Artificial Intelligence* 24(6), 494–509 (2010)
6. de Melo, C.M., Gratch, J.: Expression of Emotions Using Wrinkles, Blushing, Sweating and Tears. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsón, H.H. (eds.) *IVA 2009*. LNCS, vol. 5773, pp. 188–200. Springer, Heidelberg (2009)
7. McGregor, I.: The Sweating Reactions of the Forehead. *J. of Physiology* 116, 26–34 (1952)

8. Fautua, D., Schatz, S., Kobus, D., Spiker, V., Ross, W., Johnston, J., Nicholson, D., Reitz, E.: *Border Hunter* (OMB No. 0704-0188). Technical report, USJFC (2010)
9. Arnold, M.B.: Physiological Differentiation of Emotional States. *Psychological Review* 52(1), 35–48 (1945)
10. Drummond, P., Lance, J.: Facial Flushing and Sweating Mediated by the Sympathetic Nervous System. *Brain* 110, 793–803 (1987)
11. Williams, G., Scott-Donelan, D.: *Combat Observation and Decision-Making in Irregular and Ambiguous Conflicts (CODIAC)*. USJFC, Norfolk (2010)
12. Ortiz, E., Salcedo, J., Lackey, S., Fiorella, L., Hudson, I.: Soldier vs. Non-military Novice Performance Patterns in Remote Weapon System Research. In: *2012 Spring Simulation Multiconference–Military Modeling and Simulation Symposium*, pp. 1–6. Curran Associates, Red Hook (2012)
13. Petrides, K.V., Furnham, A.: Trait emotional intelligence: Psychometric Investigation with Reference to Established Trait Taxonomies. *European Journal of Personality* 15, 425–448 (2001)
14. Barber, D.J., Leontyev, S., Sun, B., Davis, L., Chen, J.Y.C., Nicholson, D.: The Mixed-Initiative Experimental (MIX) Testbed for Collaborative Human Robot Interactions. In: *2008 International Symposium on Collaborative Technologies and Systems*, pp. 483–489 (2008)
15. Combs, G., Luthans, F.: Diversity Training: Analysis of the Impact of Self-efficacy. *Human Resource Development Quarterly* 18(1), 91–120 (2007)
16. Tai, W.: Effects of Training Framing, General Self-efficacy and Training Motivation on Trainees' Training Effectiveness. *Personnel Review* 35(1), 51–65 (2006)
17. Salami, S.O.: Emotional Intelligence, Self-efficacy, Psychological Well-being and Students' Attitudes: Implications for Quality Education. *European J. of Ed. Studies* 2(3), 247–257 (2010)

# Author Index

- Abate, Andrea F. II-201  
Abich IV, Julian II-211  
Ahn, Sang Chul I-32  
Alfalah, Salsabeel F.M. II-3  
Altarteer, Samar II-221  
Amemiya, Tomohiro I-121, II-372  
Arizmendi, Brian J. I-129
- Ban, Yuki II-90  
Banic, Amy Ulinski I-39  
Barneche Naya, Viviana II-152  
Beckmann-Dobrev, Boris I-149  
Biocca, Frank I-333  
Boivin, Eric II-251  
Bordegoni, Monica II-291  
Boyer, James I-111  
Brahnam, Sheryl II-12  
Brandenburg, Stefan I-149  
Braz, Priscilla I-139  
Breedon, Chad II-231  
Brewer, Bambi R. II-22  
Brokaw, Elizabeth B. II-22  
Brown, Gordon II-60  
Budziszewski, Paweł II-32  
Burnett, Gary II-261
- Calhoun, Gloria II-231  
Carretta, Thomas I-111  
Caruso, Giandomenico II-291  
Cascales, Antonia II-103  
Chan, Warren II-50, II-66, II-221  
Charissis, Vassilis II-3, II-50, II-66,  
II-122, II-221  
Chella, Antonio I-253  
Chen, Chien-Hsu II-400  
Chen, Yu-Ping II-40  
Cho, Yongjoo I-58  
Choi, Woong II-409  
Clarens, Alex Rodiera I-86  
Coburn, Sarah I-3  
Conde, Miguel Ángel II-132  
Contero, Manuel II-103  
Cook, Maia B. II-320  
Corenthy, Loic I-79
- Costa, Mark R. I-333  
Courtney, Christopher G. I-129  
Crabtree, Andy II-261  
Cullen, Sean II-310
- da Costa, Rosa Maria E. Moreira II-363  
Dam, Peter I-139  
Dawson, Michael E. I-129  
de Figueiredo, José Eduardo M. II-363  
Del Pozo, Alberto II-132  
Dijk, Judith II-330  
Dill, Kevin II-310  
Dindo, Haris I-253  
Dittrich, Elisabeth I-149, II-301  
Draper, Mark II-231  
Duarte-Gonçalves, Bruno Filipe II-390  
Duran, Jaume II-113
- Edlinger, Günter II-74  
Engelke, Timo I-49
- Fahn, Chin-Shyurng II-353  
Falah, Jannat II-122  
Fiore, Stephen M. II-170  
Fonseca, David II-188  
Fröhlich, Julia I-159  
Fukuda, Yusuke I-231
- Garcia, Marcos I-79  
García-Peñalvo, Francisco José II-132  
García-Vergara, Sergio II-40  
Gauld, Dylan II-60  
Gértrudix Barrio, Manuel II-426  
Gil, Kyungwon I-343  
Gombos, Gergő I-202  
González Castro, Miguel Ramón II-271  
Goodrich, Michael A. I-267  
Graf, Holger I-13  
Green, Collin I-277
- Ha, Taejin I-343, I-359  
Hachimura, Kozaburo II-409  
Hamell, Joshua II-231  
Handa, Daiki I-23  
Hardin, Stacey E. II-142

- Harrison, David K. II-3, II-50, II-122, II-221  
 Hayes, Aleshia T. II-142  
 Hayes, Austen I-39  
 Hernández Ibáñez, Luis Antonio II-152  
 Higa, Kyota I-351  
 Hirose, Michitaka II-90  
 Hirota, Koichi I-121, II-372  
 Hirumi, Atsusi "2C" II-416  
 Holte, Michael B. II-241  
 Hoskins, Gaylor II-60  
 Houben, Mark II-330  
 Howard, Ayanna M. II-40  
 Hudson, Irwin I-211, I-388  
 Hughes, Charles E. II-142, II-162, II-426  
 Hughes, Darin E. II-162, II-426  
 Hulin, Thomas I-241  
 Hwang, Jae-In I-32
- Ikei, Yasushi I-121, II-372  
 Ishii, Hirotake I-23  
 Ishikawa, Kazuo I-175  
 Ishikawa, Masumi I-351  
 Ito, Yoko II-372
- Jaff, Lana I-39  
 Jentsch, Florian I-285, I-295, I-313  
 Jeong, Jihoon II-83  
 Johnson, Adrian S. I-169  
 Johnson-Glenberg, Mina C. II-380
- Kalar, Donald I-277  
 Kastel, Thiemo II-390  
 Kawashima, Kimiko I-175  
 Keebler, Joseph R. I-321, II-170  
 Keefe, Daniel F. II-179  
 Keil, Jens I-49  
 Kesmaecker, Marion II-390  
 Khan, Soheeb II-50, II-66  
 Kim, Dongho I-103  
 Kim, Gerard Joungyun I-73  
 Kim, Hee Jae I-65  
 Kim, Hyoung-Gon I-32  
 Kim, Ig-Jae I-32  
 Kim, Jea In I-359  
 Kim, Minyoung I-58  
 Kim, Seokhwan I-58  
 Kim, Sung Yeun I-333  
 Kimura, Asako I-221  
 Kiss, Attila I-202
- Kitazaki, Michiteru I-184  
 Ko, Heedong I-32  
 Kollreider, Alexander II-74  
 Kubota, Yosuke I-378  
 Kwon, Yeram II-83
- Lackey, Stephanie I-211, I-388  
 Laguna, Isabel II-103  
 Laidlaw, David H. II-179  
 Laurendeau, Denis I-192, II-251  
 Lebiere, Christian I-285  
 Lee, Jaedong I-73  
 Lee, Jong Weon I-65  
 Lee, Sangyong I-73  
 Lee, Sungwon II-83  
 Lee, Yi-Chia Nina II-400  
 Lévesque, Julien-Charles I-192  
 Li, Liang II-409  
 Lindgren, Robb II-162  
 Liu, Wei-Tyng II-353
- Mair, Gordon M. I-368  
 Maraj, Crystal I-211, I-388  
 Martinez, Joy I-388  
 Matuszka, Tamás I-202  
 McGhee, John II-60  
 Meas, Phe II-310  
 Mikolajczyk, Krzysztof II-390  
 Miller, Christopher II-231  
 Mokhtari, Marielle I-192, II-251  
 Moore, J. David II-66  
 Moshell, J. Michael II-162  
 Mott, Dana S. II-416  
 Murray, Jennifer II-60
- Nakanishi, Miwa I-231  
 Narducci, Fabio II-201  
 Narumi, Takuji II-90  
 Navarro, Isidro I-86  
 Negishi, Kazuno I-175  
 Nguyen Hoang, Anh I-103  
 Noh, Jin I-94  
 Nomura, Toshiyuki I-351  
 Nwakacha, Valentine II-261  
 Nyamse, Victor II-66
- Ogi, Tetsuro I-378  
 Ohta, Takashi II-436  
 Okamoto, Jun I-175  
 Oliver, James II-281

- Ortiz, Eric I-211, I-388  
 Ortner, Rupert II-74  
 Ososky, Scott I-285, I-295  
 Owen, Charles I-3  
  
 Park, Kyoung Shin I-58  
 Park, Nohyoung II-83  
 Parker, Caroline II-66  
 Parsons, Thomas D. I-129  
 Pastor, Luis I-79  
 Peng, Tao II-340  
 Peredo, Alberto II-188  
 Pérez-Cota, Manuel II-271  
 Pérez-López, David II-103  
 Perona, Pascual II-103  
 Phillips, Elizabeth I-295  
 Pitsch, Harald II-74  
 Preusche, Carsten I-241  
  
 Radkowski, Rafael II-281  
 Ram, David II-74  
 Raposo, Alberto I-139  
 Re, Guido Maria II-291  
 Rebenitsch, Lisa I-3  
 Redondo, Ernest II-188  
 Reich, Diana II-301  
 Reinerman-Jones, Lauren II-211  
 Ricciardi, Stefano II-201  
 Rizzo, Albert A. I-129  
 Rowe, Allen I-111  
 Ruff, Heath II-231  
  
 Sabbagh, Shabnam II-162  
 Sagardia, Mikel I-241  
 Sakellariou, Sophia II-50  
 Sakurai, Sho II-90  
 Salcedo, Julie I-211, I-388  
 Sánchez Riera, Albert II-188  
 San Martin, Jose I-79  
 Schaffer, Richard II-310  
 Scheutz, Matthias I-304  
 Schmitt, Michael I-49  
 Schuster, David I-313  
 Seo, Ki-Young I-58  
 Shan, Li-Ting II-400  
 Shen, Jun-wu II-340  
  
 Shi, Chung-Kon I-359  
 Shibata, Fumihisa I-221  
 Shimoda, Hiroshi I-23  
 Smallman, Harvey S. II-320  
 Smith, Dustin C. I-321, II-170  
 Stapleton, Christopher II-416  
 Stork, André I-13  
 Streefkerk, Jan Willem II-330  
 Sun, Yu I-169  
 Sung, Min-Hyuk I-32  
 Swarts, Matthew I-94  
  
 Tamura, Hideyuki I-221  
 Tanaka, Kazuki I-221  
 Tanikawa, Tomohiro II-90  
 Tanuma, Kazuhiro I-231  
 Tateyama, Yoshisuke I-378  
 Taylor, Grant II-211  
 ter Haar, Frank II-330  
 Toda, Azusa I-221  
  
 Uchiyama, Takahiro I-231  
  
 van Amerongen, Pjotr II-330  
 Venero, Peter I-111  
 Viet Tran, Hoang I-103  
 Villagrasa, Sergi II-113  
  
 Wachsmuth, Ipke I-159  
 Walters, Lori C. II-426  
 Weber, Bernhard I-241  
 Werneck, Vera Maria B. II-363  
 Wientapper, Folker I-49  
 Williams, Brian II-60  
 Wiltshire, Travis J. I-321, II-170  
 Wojtowicz, Joanna II-74  
 Woo, Woontack I-343, I-359, II-83  
 Wood, Bruce M. II-122  
 Wu, Meng-Luen II-353  
 Wu, Yun-feng II-340  
  
 Yasumoto, Masasuke II-436  
 Yi, Daqing I-267  
  
 Zhang, Ying II-340  
 Zoellner, Michael I-49