Trends in Augmentation of Human Performance 5

Seong-Whan Lee
Heinrich H. Bülthoff
Klaus-Robert Müller *Editors*

Recent Progress in Brain and Cognitive Engineering

Springer

Trends in Augmentation of Human Performance

Series editor
Vassilis Cutsuridis, Boston University, Boston, USA

Human performance enhancement refers to the augmentation of human skills, attributes and competencies through the use of technology, medicine and therapy designed to replace or increase human performance capability. In the turn of the second decade of the 21st century the convergence of the neurosciences including neuroimaging, genetics and cognitive science, neurotechnology and informatics is creating powerful tools that have the potential to significantly enhance human performance.

Three broad domains of human performance enhancement will be covered by the book series: (1) Enhancement of human performance to restore normal human capability from the disabled or dysfunctional, (2) enhancement of human performance to superhuman levels, and (3) Ethical consequences of the first two domains.

Examples of the first domain will include but not limited to restoration of sight, hearing, memory, mobility, reversing aging effects, prosthesis of limbs, brain-machine interface and genetic manipulation. Examples of the second domain will include, but not limited to enhanced memory, augmentation of vision, hearing, strength, mobility, brain-machine and machine-brain interfaces.

This series will publish works of the highest quality that advance the understanding and practical application of human performance enhancement. Research monographs, introductory and advanced level textbooks, volume editions and proceedings will be considered. Please contact the editors with your book ideas.

Finally, an advisory board of distinguished scientists from the fields on neurotechnology, neurosciences, cybernetics, cognitive sciences, psychophysics, psychiatry, genetics, neuropharmacology, neuroimaging and ethics will assist the book series editor with the selection of future book proposals. The advisory board members will also serve as reviewers of future book proposals.

More information about this series at http://www.springer.com/series/10849

Seong-Whan Lee • Heinrich H. Bülthoff
Klaus-Robert Müller
Editors

Recent Progress in Brain and Cognitive Engineering

Springer

Editors
Seong-Whan Lee
Department of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea

Klaus-Robert Müller
Machine Learning Group
Berlin Institute of Technology
Berlin, Germany

Heinrich H. Bülthoff
Department Human Perception, Cognition and Action
Max Planck Institute for Biological Cybernetics
Tübingen, Germany

ISSN 2213-1310 ISSN 2213-1329 (electronic)
Trends in Augmentation of Human Performance
ISBN 978-94-017-7238-9 ISBN 978-94-017-7239-6 (eBook)
DOI 10.1007/978-94-017-7239-6

Library of Congress Control Number: 2015950341

Springer Dordrecht Heidelberg New York London

Printed on acid-free paper

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

Preface

Brain and cognitive engineering is converging toward a better understanding of cognitive information processing in the human brain. There are first attempts to develop "humanlike" and neuromorphic artificial intelligent systems. Ultimately this could lead to better modeling, prediction and analysis of brain-related diseases.

This book provides overviews and in-depth discussions of recent studies in brain and cognitive engineering. It starts from efforts toward the understanding of perceptual, structural, and functional principles behind the high-order information processing, then considers aspects on how the brain may implement these in artificial systems that can interact intelligently with the real world. Its contributions report recent progress in brain science and cognitive engineering made by scholars from around the world.

The book covers four major topics: *Noninvasive Brain-Computer Interface*, *Cognitive- and Neural-Rehabilitation Engineering*, *Big Data Neurocomputing*, and *Early Diagnosis and Prediction of Neural Diseases*. We hope that the articles in the book will be a valuable resource for researchers in the field and also that the guiding introductory materials will be of use for the potential researchers who step into the field anew.

Thank you.

Seoul, Republic of Korea — Seong-Whan Lee
Tübingen, Germany — Heinrich H. Bülthoff
Berlin, Germany — Klaus-Robert Müller

Contents

Contributors

Regine G.M. Armann Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Department of Psychology, University of York, York, UK

Boyan Bonev Department of Statistics, University of California, Los Angeles, CA, USA

Heinrich H. Bülthoff Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Isabelle Bülthoff Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Ricardo Chavarriaga Defitech Chair in Brain-Machine Interface, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Yiyu Chen Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Kyuwan Choi BK21 Plus Global Leader Development Division in Brain Engineering, Korea University, Seoul, Republic of Korea

Siamac Fazli Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Jae-Ho Han Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Stefan Haufe Neural Engineering Group, Department of Biomedical Engineering, The City College of New York, New York, NY, USA

Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, Berlin, Germany

Bernstein Focus Neurotechnology, Berlin, Germany

Dong-Cheol Hur Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea

Iñaki Iturrate Defitech Chair in Brain-Machine Interface, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Hyun Jae Jang Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Dong-Joo Kim Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Hakseung Kim Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Il-Hwa Kim Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Jeong-Woo Kim Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Young-Tak Kim Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Jeehyun Kwag Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Ryo Kyung Lee Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Min-Ho Lee Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Seong-Whan Lee Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Robert Leeb Center for Neuroprosthetics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

José d.R. Millán Defitech Chair in Brain-Machine Interface, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Byoung-Kyong Min Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Klaus-Robert Müller Machine Learning Group, Berlin Institute of Technology, Berlin, Germany

Alessandro Nesti Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Jaeseok Park Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Sahn Woo Park Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Serafeim Perdikis Defitech Chair in Brain-Machine Interface, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Paolo Pretto Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Isabella Schlattner Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Dinggang Shen Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Ahyoung Shin Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Heung-Il Suk Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Joost Venrooij Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Christian Wallraven Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

John Williamson Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Seul-Ki Yeom Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Alan L. Yuille Department of Statistics, University of California, Los Angeles, CA, USA

Part I
Non-invasive Brain-Computer Interface

Chapter 1
Future Directions for Brain-Machine Interfacing Technology

Kyuwan Choi and Byoung-Kyong Min

Abstract Brain-machine interfaces (BMIs) are a communication technology that link humans and artificial devices through brain signals. However, development of BMI technology is currently at an impasse. First, the most commonly used BMI brain signals are principally derived from the primary sensorimotor cortices. However, these signals do not precisely reflect the diverse range of human intentions. In addition, BMI operational protocols often require users to perform mental functions that are not directly related to the goal of their task. Time is ripe to explore novel BMI control signals. Brain correlates of higher cognitive functions involved in deliberate processing of information appear as novel BMI signals with a number of appealing properties. This study suggests techniques for controlling BMIs using human higher cognitive activity in a non-invasive manner, and proposes a novel viable method based on our recent observations. Since the prefrontal cortex constitutes the highest level of the cortical hierarchy dedicated to the representation and execution of actions, these findings open the door to goal-directed intention-recognition BMI technology. This technology may help to rehabilitate or improve the cognitive performance of neurological or psychiatric patients with prefrontal dysfunctions. Cognitive BMIs should be further explored to develop practical applications and therapeutic treatments that improve the quality of life for people with sensorimotor or cognitive impairments.

Keywords Brain-machine interface • Cognition • Electroencephalography • Prefrontal cortex

K. Choi
BK21 Plus Global Leader Development Division in Brain Engineering, Korea University, Seoul, Republic of Korea

B.-K. Min (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea
e-mail: min_bk@korea.ac.kr

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_1

1.1 Overview of the Current Brain-Machine Interface (BMI) Technology

At present, the society benefits from modern technology, including various types of external devices, which are usually controlled by our hands or feet. For example, the keyboard of a computer is operated by the hand, and pedals in cars are controlled by the foot. Recently, these technologies have reached an advanced level, so that humans can control tools by other means, for example, by voice based on speech recognition technology. Because the movements of the hand, foot, and vocal chords are regulated by the brain, it would be further convenient if the external devices can be controlled directly by the brain signals (i.e., without the use of relevant muscles). This approach is known as the brain-machine interface (BMI) technology [1–5]. This technology is particularly valuable for people with disabilities who have muscle or mobility problems. Since the introduction of the first BMI paradigms [6], the approach has received growing attention. BMI technology has become one of the important topics in neuroscience and bioengineering [7–15]. This is partly because of the substantial progress in the development of algorithms and experimental settings that may assist patients with paralysis to use their brain signals to intentionally control directional motions of external devices.

BMI technology can be broadly divided into invasive or non-invasive, based on two approaches. There are several successful examples of invasive BMI [4, 16, 17]. The arm movement of a robot, with one degree of freedom, was controlled using the neural activity of the rat motor cortex [18]. Carmena et al. [19] have consistently succeeded in reconstructing the arm movement of a robot having three degrees of freedom, including grip force, by using the neural activity of the premotor cortex, primary motor cortex, and posterior parietal cortex of a monkey. In a study on humans, a computer cursor on a two-dimensional display was controlled by signals from the primary motor cortex [20]. There have also been many human studies employing non-invasive BMI methods (e.g. electroencephalography [EEG], or near-infrared spectroscopy [NIRS]) to allow people to exert volitional control over external devices [11–13, 21–28]. For example, non-invasive brain signals were used for navigating a cursor through a maze [29], controlling a wheelchair [30–34], selecting letters on a display to produce text [35–39], and performing motor imagery to guide an external direction [8, 40–47].

Several BMI studies have focused on harnessing the neural signals from specific brain areas to adapt a real-time map to the motion-control states of some external device [9, 19, 48–50]. The success of using the spike- and/or local field potential (LFP)-signals from different cortical regions for intentional control of external events has been widely recognized in both human and non-human primates [20, 48, 51–56]. In all cases, however, the brain areas for BMI signals must be selected *a priori*, often based on the known properties of the area and its computational strengths for motor or cognitive control (or both). As with the spike- and LFP-signals, methods using event related potentials (ERPs) [57] can also provide an accurate and reliable control of external events when the brain regions are

determined *a priori* and when the signals from the appropriate regions are used and co-adapted in real-time with the control states of the external device.

Accordingly, EEG is an ideal modality for the implementation of BMI technology in a larger population. It is portable, low cost, and lacks the inevitable risks associated with invasive approaches [3, 58]. In general, the EEG-based BMI paradigms can be subdivided into two types: (1) externally stimulated paradigms such as steady-state visual evoked potential (SSVEP) [59] and P300 [37, 60], and (2) internally induced paradigms such as motor imagery [3, 12, 45, 61–63]. For these paradigms, the brain activity of interest is detected in the occipital, parietal, and central cortex, respectively. It is noteworthy that the prefrontal cortex has been largely ignored in the BMI research. This is an important oversight because certain goal-directed cognitive processes are regulated by the prefrontal cortex, which involves cognitive action [64] and goal-directed behaviors [64, 65]. Moreover, prefrontal cortical activity can be an alternative in the case of patients with damage to other cortical areas, or when BMI signals derived therein are not reliable. Therefore, these prefrontal brain correlates may open a new avenue for the development of goal-directed intention-recognition BMI technology. This is in part because the prefrontal cortex constitutes the highest stage of the cortical hierarchy dedicated to represent actions [66].

1.2 Cognitive BMI

The brain correlates of higher cognitive functions that are involved in the deliberate processing of information appear as novel BMI signals. Moreover, higher cognitive functions entail direct planning to achieve goals, independent of the sensory modalities that convey the status of the task under consideration. For instance, the prefrontal cortex activity estimated from the EEG signals reflects the goal-directed intentional processes for a simple binary control task presented on a computer monitor [67]. In this case, subjects were asked to choose either a left or right direction, based on the direction of a previously presented cue (an arrow), and then to directly control a bar on the monitor to initiate movement in the intended direction. Although the subjects simply imagined the cued direction, which is more intuitive than task-irrelevant motor imagery, their goal direction was reliably decoded from the prefrontal cortex activity.

As an example of future directions in BMI technology, we describe a sample of cognitive BMI studies using the prefrontal signal paradigm. In this paradigm, a simple binary control task combines the visual feedback from the subject's real-time performance with motor imagery in a closed loop to engage the subject in a co-adaptation process (between the internal cortical signal and the external performance). Motor imagery enables us to provide immediate on-line feedback of the performance based on a 125-ms sliding window that permits spontaneous co-adaptation between the neural signal and the instantaneous visual feedback of the desired external cursor direction. This is possible because the subjects are not

consciously aware of the processing and transmission delays. Although we cannot fully parameterize the internal motor imagery, we can estimate the neural activation across all 64 channels of EEG, and the features that maximize accuracy can be automatically extracted. In this way, instead of *a priori* harnessing of the neural signal from predefined brain areas, we can set a Bayesian sparse probit classifying algorithm to automatically select the regions with highest activation for solving the specific intentional task.

Fourteen subjects with college education (mean age = 29.0 ± 7.0 years; 7 males, 7 females) participated in this experiment. The experiment consisted of two phases: (1) training, open-loop (no feedback from the performance) and, (2) training, closed-loop (with feedback from the performance). The training phase was guided by a visual input with the stimulus being an arrow on the screen monitor. The arrow pointed to one of two locations, right or left, to indicate to the subject the direction they had to imagine and intend to move the cursor. In the open-loop version of the training phase, the subject imagined the instructed direction but received no feedback concerning the outcome of this imagination exercise. Figure 1.1a shows the schematics of the epochs of the open-loop training phase. Two seconds of fixation were followed by 4 s of display of the direction that the subject had to imagine (left or right); direction alternatives were randomly presented. The last epoch was the "resting phase" lasting 3–4 s; total trial length was 9–10 s. The EEG activity during this open-loop training phase was used as a baseline for late adaptation of the signal and to correctly modulate the performance outcome. There were 7 blocks of 30 trials each during this open-loop training phase.

In the present study, source currents over 2,240 vertices were estimated for the EEG signals [68–71] from 64 channels through a hierarchical Bayesian method [72–74] that can effectively incorporate both structural and functional MRI data. Accordingly, in our study, the variance of the source current at each source location is considered as an unknown parameter and is estimated from the observed EEG data. Prior information is also incorporated using the hierarchical Bayesian method [75]. The fMRI information was imposed as prior information onto the variance distribution (rather than the variance itself) to give a soft constraint on the variance. After estimating the source currents from the EEG signals, a sparse probit classifier [76, 77] classifies the left- and right-hand motor imagery. By using the source currents estimated from the EEG signals as features, the number of features can be increased, and among the increased features, the sparse probit classifier automatically selects only the useful features for classification. Therefore, the method is very robust against the over-fitting, and the accuracy of classification can be improved. Furthermore, by using the source currents as features instead of the response variance, the subjects receive immediate feedback of their intentions rather than an estimate of a by-product signal. For these reasons, the source currents were used as features in this study although the common spatial pattern (CSP) without source localization has commonly been used as features for motor-imagery BMI paradigms [27]. This is attributed to the concept that the sources cannot contain more information than in the scalp electrodes, because the inverse problem is solved with a linear method [78].

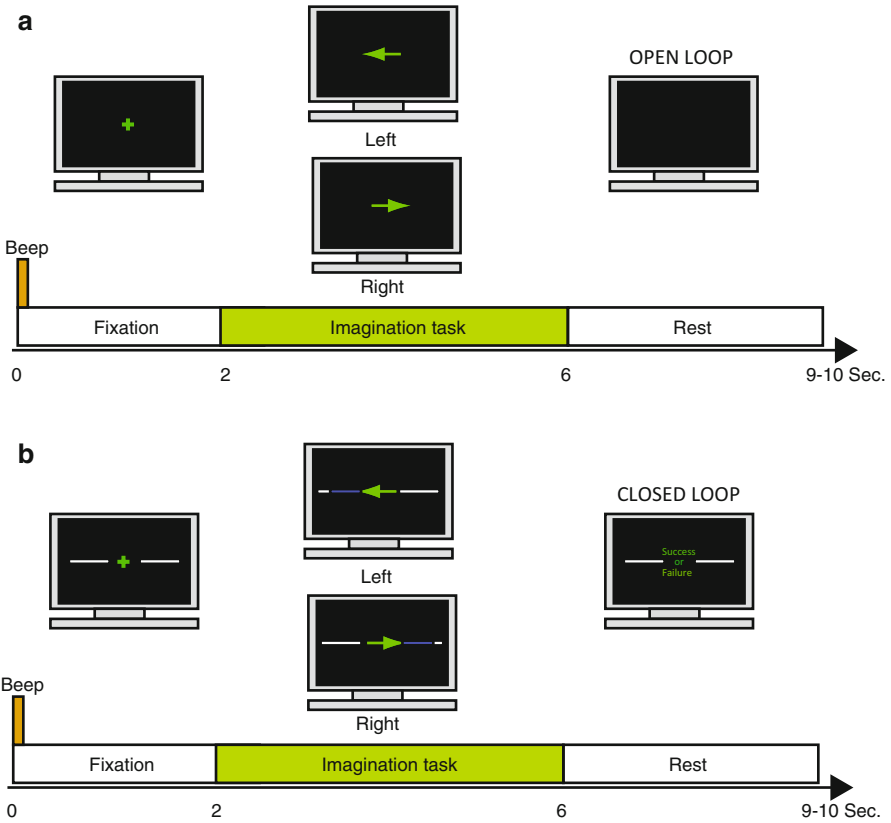

Fig. 1.1 Experimental design for BMI cursor control: (**a**) Open-loop version of the training phase where no performance feedback is provided. Epochs of one trial include fixation on a cross (2 s), imagination task (4 s), and resting phase (3–4 s). (**b**) The closed-loop version of the training phase (brain control) provides the subject with a performance feedback for trying to correctly move the cursor according to the instructed direction, by thinking about that direction. The EEG signal is changed from trial to trial according to the classifier algorithm to accurately control the motion of the cursor in real-time and in compliance with the instructed direction of the *green arrow* displayed on the monitor. The subject receives instantaneous visual feedback from the real-time performance in each trial, by shifting the *blue progressing bar* in the direction of the *arrow*. At trial completion, the subject views a success or failure message on the monitor. At the end of each of the 7 blocks (30 trials per block), the subject receives feedback on their success rate according to the percentage of correct trials

In the training phase using the closed-loop task (Fig. 1.1), subjects receive instantaneous performance feedback in the form of a blue progressing bar that moves (or fails to move) in the direction instructed by the arrow. A learning progression occurs because of the co-adaptation between the EEG real-time signal and the subject's performance (via the visual feedback of the real-time moving bar). The subject receives feedback from this learning progression after each trial. The visual feedback, from the animation of the bar moving left or right, informs the subject of success or failure and is accompanied by the text string "SUCCESS" or

"FAILURE" (Fig. 1.1b, last panel). Seven blocks of 30 trials are used to train the subject's EEG signal in the closed-loop task with bar feedback. At the end of each block of trials, the subject is also informed of their percentage of correct trials, which assists in tracking the rate of learning. As with the open-loop phase, the closed-loop training phase lasts for 35 min.

Figure 1.2 shows the overall configuration of the real-time BMI system used in this first study. There are five main steps: (1) Registering the EEG signal with the cortical surface model using the 3-layer head model (brain, skull, and scalp) [79] to estimate the lead field matrix G [80, 81] (to obtain the forward map from the current sources to the 64 sites of the electrodes); (2) Estimating the forward map in the open-loop training phase to establish a baseline estimate of the variance of the current, which will be changed next in the closed-loop phase of training; (3) Estimating the inverse filter to generate the cortical activity provided to the EEG signal; (4) Selecting the cortical features (regions) that lead to successful classification of the desired direction and correct shift of the cursor; (5) Updating the filters and classifiers continuously (in real-time) using a sliding window of 125 ms to maximize

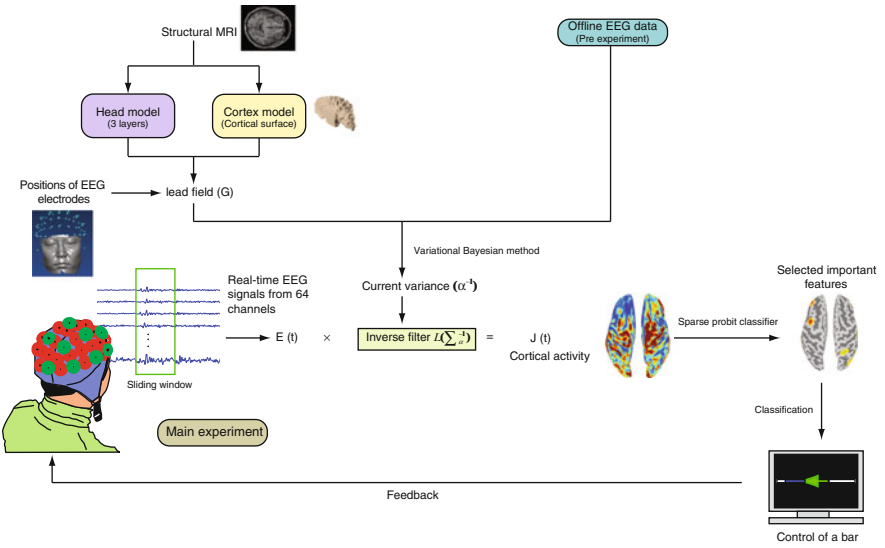

Fig. 1.2 General concept of BMIs using the EEG data. Baseline EEG data analyzed off-line (from the open-loop segment of the training phase) was used in combination with the head model data from the subject. To build the head model data, the structural MRI template obtained previously is used in combination with the head model (3 layers) and the cortex model. These data are fit to the actual physical positions and orientations of the electrodes. The real-time data is processed using variational Bayesian methods with a sliding window of 125 ms and passed through an inverse filter to obtain the cortical activation. A sparse probit classifier is used to select the important features of the data and highlight the cortical regions where maximal activation is evoked by the closed-loop training task. The bar on the screen moves in real-time, as the subject thinks left or right, according to the minimization of the error between the desired direction of the *arrow* on the monitor and direction provided by the control algorithm. Instantaneous visual feedback is provided to the subject from the real-time performance and the percentage of correct trials is displayed at the end of each block

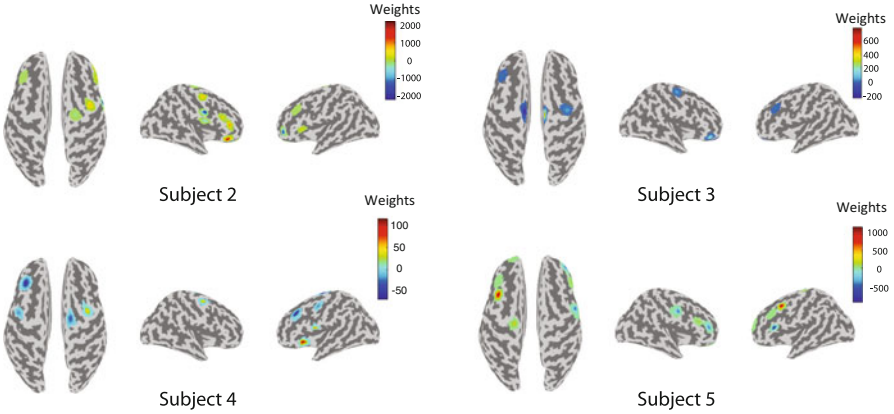

Fig. 1.3 Important task-relevant features selected by the classifier after completion of the feedback training

the probability of correctly controlling the motion of the bar in the desired direction. Steps 1 and 2 are performed offline in the open-loop training phase to establish the baseline information for each subject. Step 3 is performed once in the offline open-loop phase, and then Steps 3–5 are performed in the closed-loop training to change the filters and classifiers appropriately. The resulting signal can be tracked on the cortical surface as the subject learns the task in real-time. In addition, the evolution of brain activity can be tracked after a certain number of trials, or plotted as an averaged activation across all 210 trials (7 blocks, 30 trials per block).

Figure 1.3 shows the features selected by the sparse probit classifier after the subjects completed the fourth session. The features selected by the classifier are not most meaningful but patterns [57, 82, 83]. After feedback training, features are located mainly in the prefrontal cortex. The processing of the motor-related information in the brain is thought to follow a progression: first, the urge to move the arm is reflected in the premotor cortex; then the signal shifts to the primary motor cortex via the supplementary motor area. The primary motor cortex, assumed to be the final output portion of the motor-related signals in the brain, sends the signal to the muscles through alpha motor neurons of the spinal cord, thus resulting in the physical movement of the arm [84]. In our study, however, once the subjects perform well above the chance level, the prefrontal cortex is automatically recruited. The prefrontal cortex is involved in many aspects of planning and cognitive control. The present results imply that a more abstract representation of the motor imagery signal exists, possibly indicating a neural correlate for intentionality in this region. Figure 1.4 shows the Brodmann area (BA) 6 engaged in the directional control task and logically poses the question of whether this signal precedes the motor command. Figure 1.4 shows the important features selected by the sparse probit classifier over the selected time interval and while one subject (Subject 3) performed the motor imagery task. Between 0 and 0.2 s, features in the prefrontal cortex were selected, whereas from 0.2 to 4 s, features in the BA6 were chosen. This result

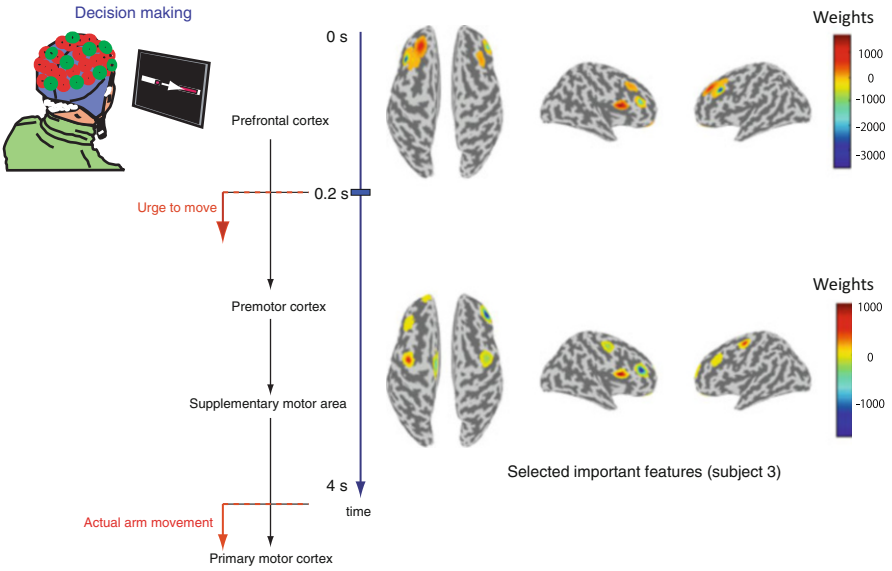

Fig. 1.4 Important features selected by the sparse probit classifier over the selected time interval while Subject 3 performs the motor imagery task

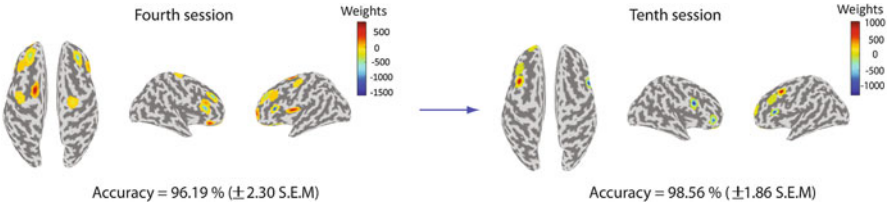

Fig. 1.5 Results for long-term training (after the 10th session) in Subject 1

suggests that the subjects may use the prefrontal cortex as part of a planning stage for impending motor decisions prior to engaging BA4 (primary motor cortex) in the output execution stage. This raises the question: what will the result be after training the subject for a long period?

Feedback training was provided continuously to one subject (Subject 1) over the 4th session, in which he showed almost perfect performance (see Fig. 1.5). Subject 1 was then trained over an extended period, completing 10 sessions. The accuracy in the 10th session was almost the same as that in the 4th session, that is, there was no apparent improvement in accuracy. However, in the 4th session, the subject mainly used the BA6 (supplementary motor area) to control the bar's direction. In contrast, long-term training decreased the BA6 activation. In addition, all subjects reported that as they mastered the mental control of the cursor they thought less of the direction and felt a spontaneous, rather than a deliberate performance. The subjects unanimously reported that they moved the bar left or right intentionally but without explicit thoughts of direction. It is possible that this experience is

accompanied by the recruitment of the prefrontal cortex that we quantified here. The lack of recruitment of the BA4 (primary motor cortex) suggests that a more abstract representation of motion may exist in the prefrontal cortex, associated with BA6, and then towards more frontal regions upon further training.

Ojakangas et al. [85] reported that the human prefrontal/premotor cortical neurons could provide information about the planning of movement and decision-making, sufficient to decode the planned direction of movement. Our results are consistent with their observations. In addition, we have also tested blindfolded subjects by providing auditory commands and found comparable levels of performance accuracy. This suggests that a true abstract amodal signal exists in the prefrontal cortex, and that this signal is useful, ultimately, to spontaneously direct our thoughts.

Figure 1.6 shows the estimated cortical activities when the subjects performed a directional imagination task for 35 min, that is, when controlling a bar on the monitor by imagining its direction and getting feedback from the movement of the bar. As shown in the figure, the prefrontal cortices of all the subjects are activated. Figure 1.7a shows the results when the subjects simply move their eyes to the left

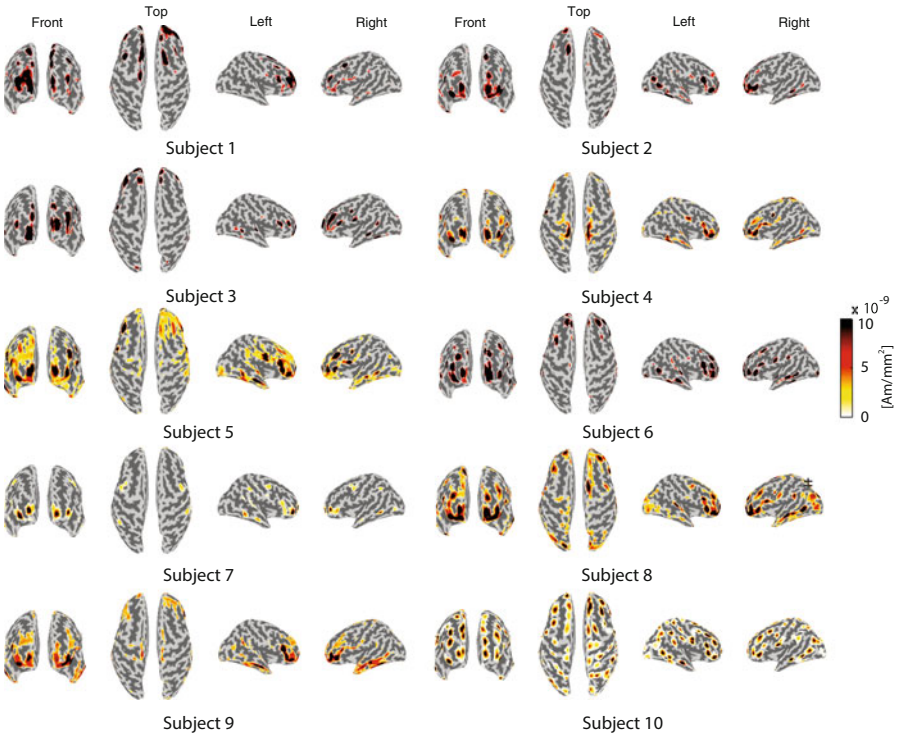

Fig. 1.6 Patterns of cortical activation in 10 subjects. Frontal, top, right, and left views of the brain (registered with activation according to the output of the sparse probit classifier), showing the most useful features for maximizing the probability of successfully moving the cursor in the intended direction. The activation of the corresponding areas are highlighted (Am/mm^2), obtained by averaging across the 4 s testing phase

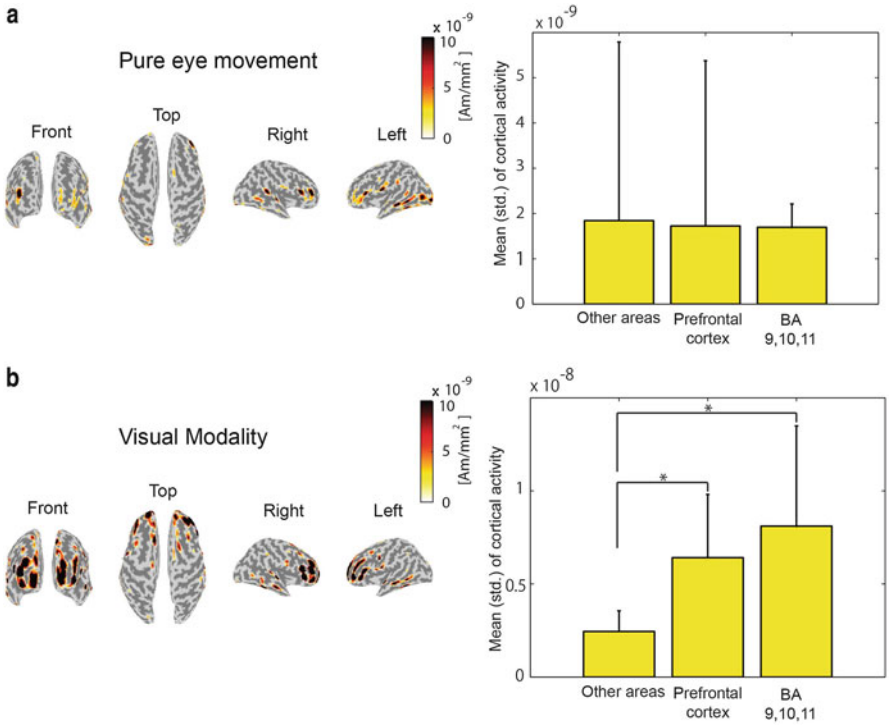

Fig. 1.7 Averaged activation patterns across subjects. (**a**) Control test for pure eye movements reveals no significant differences in recruitment of the prefrontal cortical areas in relation to the rest of the brain. (**b**) On average, across subjects, the visually driven control task recruits the prefrontal cortex and particularly, the Brodmann areas 9, 10, and 11 with patterns of activation significantly than the rest of the brain (*, $p < 0.001$)

and right. The prefrontal cortex (BA9, 10, and 11) and other areas have almost the same level of task-relevant activity. Figure 1.7b shows the average of the estimated cortical activities for 14 subjects performing the directional imagination task. The average value for the prefrontal cortex is 2.7 times higher than that for other areas. In particular, there is strong activation in the BA9, 10, and 11 located in the center of the prefrontal cortex. Taken together, these observations may provide prospective evidence that the prefrontal cortex activity can be used as a potent cognitive signal for future BMI technology.

1.3 The Future of BMI

There is great value in utilizing the knowledge accumulated in the field of neuroscience for the general welfare of humankind. Presumably, the world in which we can control a machine or a computer simply by thought might be very near. Indeed, current BMI technology can already play a role as supportive devices for

Fig. 1.8 Example of an EEG-based BMI: A wheelchair system controlled by EEG signals [31]

people in wheelchair (see Fig. 1.8). Future BMI technologies may even play an essential role in many other areas (e.g., home, office, or entertainment) with the potential to enrich our lives [5, 86–88].

However, current BMI technology has several limitations and still involves relatively primitive commands to control external devices. Of course, the extraction of binary information from the EEG signals across the left and right hemispheres is less complex. Although computer technology has made amazing progress using binary information, BMI technology still has a long way to utilize such binary information to clearly reflect our complicated mental world in real-time. Currently, it is difficult to clearly detect our various thoughts for the purpose of intentional control of BMIs. Even the EEG modality, which is the most popular for implementing BMI, has inherent limitations in that its spatial resolution is poor. To overcome this limitation, "hybrid BMI" has been proposed; it combines multiple brain imaging modalities [89].

In the future, if one's brain signals can be delivered to another person's brain through wireless communication, it could broaden the BMI field. For example, if low-intensity focused-ultrasound sonication (a non-invasive, machine-to-brain interfacing technique) is verified for its feasibility and safety, it could be used to study the transfer of signals from one brain to another in combination with the BMI technology [Fig. 1.9; 90]. The recent development of this low-intensity focused-ultrasound technology has increased the possibility of changing the brain function in a non-invasive manner [91, 92]. Clearly, there are social and ethical issues that we have to considerably concern, as BMI technology progresses to a level in the future in which we use it in daily life.

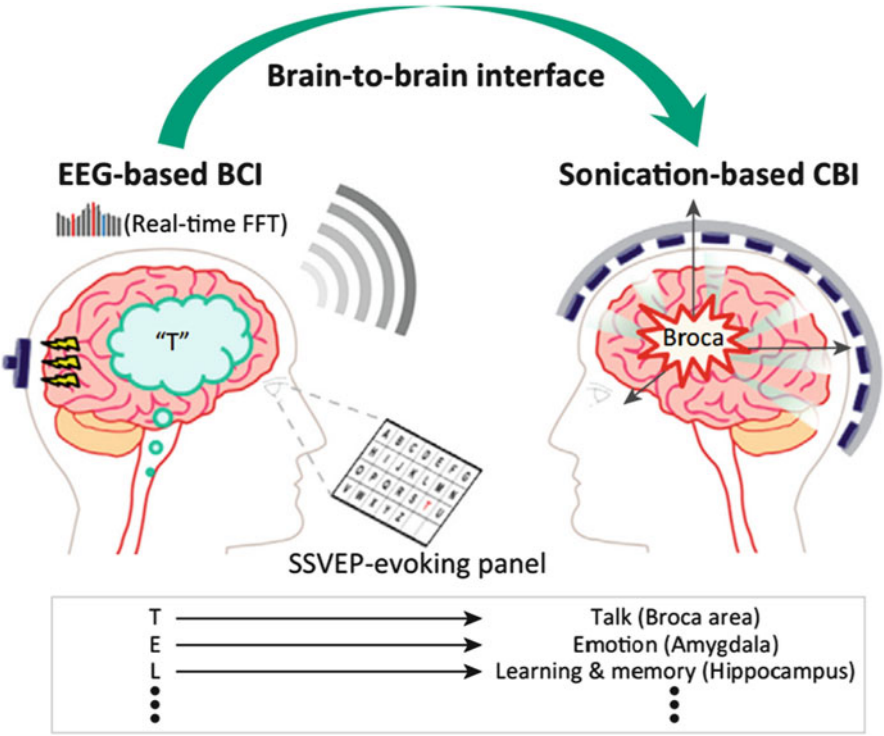

Fig. 1.9 Schematic representation of the brain-to-brain interfacing technology [90]

Acknowledgements This work was supported by the Basic Science Research Program (grant number 2012R1A1A1038358) and the BK21 Plus program, funded by the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea. The authors acknowledge the use of text from the prior publications [3, 67, 90–92]. KC and BKM thank their co-authors for allowing them to use materials from prior joint publications [67, 90].

References

1. Dornhege G, Millán JDR, Hinterberger T, McFarland D, Müller K-R (2007) Towards brain-computer interfacing. The MIT Press, Cambridge, MA
2. Wolpaw J, Wolpaw EW (2012) Brain-computer interfaces: principles and practice. Oxford University Press, New York
3. Min BK, Marzelli MJ, Yoo SS (2010) Neuroimaging-based approaches in the brain-computer interface. Trends Biotechnol 28:552–560
4. Choi K, Hirose H, Sakurai Y, Iijima T, Koike Y (2009) Prediction of arm trajectory from the neural activities of the primary motor cortex with modular connectionist architecture. Neural Netw 22:1214–1223
5. Millán JDR, Rupp R, Muller-Putz GR, Murray-Smith R, Giugliemma C, Tangermann M et al (2010) Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. Front Neurosci, 07 September 2010. http://dk.doi.org/10.3389/mins.2010.00161

6. Vidal JJ (1973) Toward direct brain-computer communication. Annu Rev Biophys Bioeng 2:157–180
7. Contreras-Vidal JL, Bradberry TJ (2011) Design principles for noninvasive brain-machine interfaces. Conf Proc IEEE Eng Med Biol Soc 2011:4223–4226
8. Contreras-Vidal JL, Bradberry TJ, Agashe H (2010) Movement decoding from noninvasive neural signals. Conf Proc IEEE Eng Med Biol Soc 2010:2825–2828
9. Lebedev MA, Carmena JM, O'Doherty JE, Zacksenhouse M, Henriquez CS, Principe JC et al (2005) Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. J Neurosci 25:4681–4693
10. Millán JR, Renkens F, Mourino J, Gerstner W (2014) Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans Biomed Eng 51:1026–1033
11. Wolpaw JR, McFarland DJ (1994) Multichannel EEG-based brain-computer communication. Electroencephalogr Clin Neurophysiol 90:444–449
12. Blankertz B, Dornhege G, Krauledat M, Muller KR, Curio G (2007) The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. Neuroimage 37:539–550
13. Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kubler A et al (1999) A spelling device for the paralysed. Nature 398:297–298
14. Scherer R, Muller GR, Neuper C, Graimann B, Pfurtscheller G (2004) An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. IEEE Trans Biomed Eng 51:979–984
15. Choi K (2013) Reconstructing for joint angles on the shoulder and elbow from non-invasive electroencephalographic signals through electromyography. Front Neurosci, 24 October 2013. http://dx.doi.org/10.3389/mins.2013.00190
16. Andersen RA, Kellis S, Klaes C, Aflalo T (2014) Toward more versatile and intuitive cortical brain-machine interfaces. Curr Biol 24:R885–R897
17. Chase SM, Kass RE, Schwartz AB (2012) Behavioral and neural correlates of visuomotor adaptation observed through a brain-computer interface in primary motor cortex. J Neurophysiol 108:624–644
18. Chapin JK, Moxon KA, Markowitz RS, Nicolelis MA (1999) Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nat Neurosci 2:664–670
19. Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF et al (2003) Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol 1, E42
20. Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH et al (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442:164–171
21. Fazli S, Mehnert J, Steinbrink J, Curio G, Villringer A, Muller KR et al (2012) Enhanced performance by a hybrid NIRS-EEG brain computer interface. Neuroimage 59:519–529
22. Huggins JE, Levine SP, BeMent SL, Kushwaha RK, Schuh LA, Passaro EA et al (1999) Detection of event-related potentials for development of a direct brain interface. J Clin Neurophysiol 16:448–455
23. Levine SP, Huggins JE, BeMent SL, Kushwaha RK, Schuh LA, Passaro EA et al (1999) Identification of electrocorticogram patterns as the basis for a direct brain interface. J Clin Neurophysiol 16:439–447
24. Levine SP, Huggins JE, BeMent SL, Kushwaha RK, Schuh LA, Rohde MM et al (2000) A direct brain interface based on event-related potentials. IEEE Trans Rehabil Eng 8:180–185
25. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain-computer interfaces for communication and control. Clin Neurophysiol 113:767–791
26. Wolpaw JR, McFarland DJ, Neat GW, Forneris CA (1991) An EEG-based brain-computer interface for cursor control. Electroencephalogr Clin Neurophysiol 78:252–259
27. Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8:441–446
28. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 25:41–56

29. Huang D, Lin P, Fei DY, Chen X, Bai O (2009) EEG-based online two-dimensional cursor control. Conf Proc IEEE Eng Med Biol Soc 2009:4547–4550
30. Carlson TE, Millán JR (2013) Brain-controlled wheelchairs: a robotic architecture. IEEE Rob Autom Mag 20:65–73
31. Choi K (2012) Control of a vehicle with EEG signals in real-time and system evaluation. Eur J Appl Physiol 112:755–766
32. Galan F, Nuttin M, Lew E, Ferrez PW, Vanacker G, Philips J et al (2008) A brain-actuated wheelchair: asynchronous and non-invasive brain-computer interfaces for continuous control of robots. Clin Neurophysiol 119:2159–2169
33. Choi K, Sato M, Koike Y (2006) Consideration of the embodiment of a new, human-centered interface. IEICE Trans Inf Syst E89d:1826–1833
34. Millán JR, Galan F, Vanhooydonck D, Lew E, Philips J, Nuttin M (2009) Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. Conf Proc IEEE Eng Med Biol Soc 2009:3361–3364
35. Rivet B, Cecotti H, Phlypo R, Bertrand O, Maby E, Mattout J (2010) EEG sensor selection by sparse spatial filtering in P300 speller brain-computer interface. Conf Proc IEEE Eng Med Biol Soc 2010:5379–5382
36. Blankertz B, Dornhege G, Krauledat M, Schroder M, Williamson J, Murray-Smith R et al (2006) The Berlin brain-computer interface presents the novel mental typewriter Hex-o-Spell. Available: http://eprints.nuim.ie/1786/1/HamiltonBerlinBCI.pdf. Item Resolution URL
37. Donchin E, Coles MGH (1988) Is the P300 component a manifestation of context updating. Behav Brain Sci 11:357–374
38. Millán JR (2003) Adaptive brain interfaces. Commun Acm 46:74–80
39. Perdikis S, Leeb R, Williamson J, Ramsay A, Tavella M, Desideri L et al (2014) Clinical evaluation of BrainTree, a motor imagery hybrid BCI speller. J Neural Eng 11:036003
40. Bell CJ, Shenoy P, Chalodhorn R, Rao RP (2008) Control of a humanoid robot by a noninvasive brain-computer interface in humans. J Neural Eng 5:214–220
41. Kamousi B, Amini AN, He B (2007) Classification of motor imagery by means of cortical current density estimation and Von Neumann entropy. J Neural Eng 4:17–25
42. Onose G, Grozea C, Anghelescu A, Daia C, Sinescu CJ, Ciurea AV et al (2012) On the feasibility of using motor imagery EEG-based brain-computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. Spinal Cord 50:599–608
43. Qin L, Ding L, He B (2004) Motor imagery classification by means of source analysis for brain-computer interface applications. J Neural Eng 1:135–141
44. Shan H, Yuan H, Zhu S, He B (2012) EEG-based motor imagery classification accuracy improves with gradually increased channel number. Conf Proc IEEE Eng Med Biol Soc 2012:1695–1698
45. Blankertz B, Dornhege G, Schafer C, Krepki R, Kohlmorgen J, Muller KR et al (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans Neural Syst Rehabil Eng 11:127–131
46. Choi K (2013) Electroencephalography (EEG)-based neurofeedback training for brain-computer interface (BCI). Exp Brain Res 231:351–365
47. Lew EYL, Chavarriaga R, Silvoni S, Millan JD (2014) Single trial prediction of self-paced reaching directions from EEG signals. Front Neurosci, 01 August 2014. http://dx.doi.org/10.3389/mins.2013.00222
48. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP (2002) Instant neural control of a movement signal. Nature 416:141–142
49. Taylor DM, Tillery SI, Schwartz AB (2002) Direct cortical control of 3D neuroprosthetic devices. Science 296:1829–1832
50. Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK et al (2000) Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 408:361–365

51. Andersen RA, Burdick JW, Musallam S, Pesaran B, Cham JG (2004) Cognitive neural prosthetics. Trends Cogn Sci 8:486–493
52. Hatsopoulos N, Joshi J, O'Leary JG (2004) Decoding continuous and discrete motor behaviors using motor and premotor cortical ensembles. J Neurophysiol 92:1165–1174
53. Hauschild M, Mulliken GH, Fineman I, Loeb GE, Andersen RA (2012) Cognitive signals for brain-machine interfaces in posterior parietal cortex include continuous 3D trajectory commands. Proc Natl Acad Sci U S A 109:17075–17080
54. Mulliken GH, Musallam S, Andersen RA (2008) Decoding trajectories from posterior parietal cortex ensembles. J Neurosci 28:12913–12926
55. Mulliken GH, Musallam S, Andersen RA (2008) Forward estimation of movement state in posterior parietal cortex. Proc Natl Acad Sci U S A 105:8170–8177
56. Pohlmeyer EA, Solla SA, Perreault EJ, Miller LE (2007) Prediction of upper limb muscle activity from motor cortical discharge during reaching. J Neural Eng 4:369–379
57. Blankertz B, Lemm S, Treder M, Haufe S, Muller KR (2011) Single-trial analysis and classification of ERP components–a tutorial. Neuroimage 56:814–825
58. Millán JR, Carmena JM (2010) Invasive or noninvasive: understanding brain-machine interface technology. IEEE Eng Med Biol Mag 29:16–22
59. Muller-Putz GR, Pfurtscheller G (2008) Control of an electrical prosthesis with an SSVEP-based BCI. IEEE Trans Biomed Eng 55:361–364
60. He ZS, Cichocki A, Xie SL, Choi K (2010) Detecting the number of clusters in n-way probabilistic clustering. IEEE Trans Pattern Anal Mach Intell 32:2006–2021
61. Pfurtscheller G, Brunner C, Schlogl A, Lopes da Silva FH (2006) Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage 31:153–159
62. Pfurtscheller G, Neuper C (1997) Motor imagery activates primary sensorimotor area in humans. Neurosci Lett 239:65–68
63. Millán JDR, Ferrez PW, Galán F, Lew E, Chavarriaga R (2008) Non-invasive brain-machine interaction. Int J Pattern Recognit Artif Intell 22:959–972
64. Fuster JM (1989) The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe, 2nd edn. Raven, New York
65. Wood JN, Grafman J (2003) Human prefrontal cortex: processing and representational perspectives. Nat Rev Neurosci 4:139–147
66. Fuster JM (1997) The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe, 3rd edn. Lippincott-Raven, Philadelphia
67. Choi K, Torres EB (2013) Intentional signal in prefrontal cortex generalizes across different sensory modalities. J Neurophysiol 67. doi:10.1152/jn.00505.2013
68. Grave de Peralta Menendez R, Gonzalez Andino S, Perez L, Ferrez PW, Millán JdR (2005) Non-invasive estimation of local field potentials for neuroprosthesis control. Available: http://infoscience.epfl.ch/record/83251. Item Resolution URL
69. Cincotti F, Mattia D, Aloise F, Bufalari S, Astolfi L, De Vico Fallani F et al (2008) High-resolution EEG techniques for brain-computer interface applications. J Neurosci Methods 167:31–42
70. Goel MK, Chavarriaga R, Millan Jdel R (2014) Comparing BCI performance using scalp EEG- and inverse solution-based features. In: 6th international brain-computer interface conference 2014, presented at the comparing BCI performance using scalp EEG- and inverse solution-based features, 16–19 September 2014, Graz, Austria
71. Goel MK, Chavarriaga R, Millan Jdel R (2011) Cortical vs surface EEG for event related potentials-based brain-computer interfaces. In: Presented at the 5th international IEEE EMBS conference on neural engineering, 27–30 September 2011, Berlin, Germany
72. Batenburg P, O'hagan A, Veenstra R (1994) Bayesian discovery sampling in financial auditing: a hierarchical prior model for substantive test sample sizes. The Statistician 43:99–110
73. Sato MA, Yoshioka T, Kajihara S, Toyama K, Goda N, Doya K et al (2004) Hierarchical Bayesian estimation for MEG inverse problem. Neuroimage 23:806–826

74. Yamashita O, Sato MA, Yoshioka T, Tong F, Kamitani Y (2008) Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. Neuroimage 42:1414–1429
75. Neal R (1996) Bayesian learning for neural networks. Springer, New York
76. Balakrishnan S, Madigan D (2008) Algorithms for sparse linear classifiers in the massive data setting. J Mach Learn Res 9:313–337
77. Krishnapuram B, Carin L, Figueiredo MA, Hartemink AJ (2005) Sparse multinomial logistic regression: fast algorithms and generalization bounds. IEEE Trans Pattern Anal Mach Intell 27:957–968
78. Gavaret M, Trebuchon A, Bartolomei F, Marquis P, McGonigal A, Wendling F et al (2009) Source localization of scalp-EEG interictal spikes in posterior cortex epilepsies investigated by HR-EEG and SEEG. Epilepsia 50:276–289
79. Bekey GA (2005) Autonomous robots: from biological inspiration to implementation and control. MIT Press, Cambridge, MA
80. Stiefelhagen R, Ekenel HK, Fugen C, Gieselmann P, Holzapfel H, Kraft F et al (2007) Enabling multimodal human-robot interaction for the Karlsruhe humanoid robot. IEEE Trans Robotics 23:840–851
81. Choi K (2014) Electroencephalography-based real-time cortical monitoring system that uses hierarchical Bayesian estimations for the brain-machine interface. J Clin Neurophysiol 31:218
82. Lemm S, Blankertz B, Dickhaus T, Muller KR (2011) Introduction to machine learning for brain imaging. Neuroimage 56:387–399
83. Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B et al (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87:96–110
84. Fried I, Katz A, McCarthy G, Sass KJ, Williamson P, Spencer SS et al (1991) Functional organization of human supplementary motor cortex studied by electrical stimulation. J Neurosci 11:3656–3666
85. Ojakangas CL, Shaikhouni A, Friehs GM, Caplan AH, Serruya MD, Saleh M et al (2006) Decoding movement intent from human premotor cortex neurons for neural prosthetic applications. J Clin Neurophysiol 23:577–584
86. Tan DS, Nijholt A (2010) Brain-computer interfaces: applying our minds to human-computer interaction. Springer, London
87. Blankertz B, Tangermann M, Vidaurre C, Fazli S, Sannelli C, Haufe S et al (2010) The Berlin brain-computer interface: non-medical uses of BCI technology. Front Neurosci 4:198
88. Muller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 167:82–90
89. Muller-Putz GR, Breitwieser C, Cincotti F, Leeb R, Schreuder M, Leotta F et al (2011) Tools for brain-computer interaction: a general concept for a hybrid BCI. Front Neuroinform 5:30
90. Min BK, Müller KR (2014) Electroencephalography/sonication-mediated human brain-brain interfacing technology. Trends Biotechnol 32:345–346
91. Min BK, Bystritsky A, Jung KI, Fischer K, Zhang Y, Maeng LS et al (2011) Focused ultrasound-mediated suppression of chemically-induced acute epileptic EEG activity. BMC Neurosci 12:23
92. Min BK, Yang PS, Bohlk M, Park SS, Vago DR, Maher TJ et al (2011) Focused ultrasound modulates the level of cortical neurotransmitters: potential as a new functional brain mapping technique. Int J Imaging Syst Technol 21:232–240

Chapter 2
Brain-Computer Interface for Smart Vehicle: Detection of Braking Intention During Simulated Driving

Jeong-Woo Kim*, Il-Hwa Kim*, Stefan Haufe, and Seong-Whan Lee

Abstract It is most essential to stop a vehicle in time for assuring a driver's safety. In this study, a simulated driving environment was implemented to study the neural correlation of braking intention in diverse driving situations. We further investigated to what extent these neural correlates can be used to detect a participant's braking intention prior to the behavioral response. A feature combination method was proposed for the enhancement of detection performance and additional classification of emergency braking triggered by stimuli and voluntary braking. It consists of event-related potential (ERP), readiness potential (RP), and event-related desynchronization (ERD) features. Fifteen participants drove a virtual vehicle and were exposed to the diversified traffic situations in the constructed simulator framework, while technical signals (i.e., gas pedal and brake pedal), electroencephalogram (EEG) and electromyogram (EMG) signals were measured. After that, the neural correlation of the measured signals was analyzed. The proposed framework shows excellent detection performance for various kinds of driver's braking intention. Our study suggests that a driver's braking intention is characterized by specific neural patterns of sensory perception and processing, as well as motor preparation and execution, which can be utilized by smart vehicle technology.

*Author contributed equally with all other contributors.

J.-W. Kim • I.-H. Kim • S.-W. Lee (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: jw_kim@korea.ac.kr; ih_kim@korea.ac.kr; sw.lee@korea.ac.kr

S. Haufe
Neural Engineering Group, Department of Biomedical Engineering, The City College of New York, 140 Street & Convent Avenue, New York, NY 10031, USA

Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, Marchstr. 23, 10587 Berlin, Germany

Bernstein Focus Neurotechnology, Berlin, Germany
e-mail: stefan.haufe@tu-berlin.de

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_2

Keywords Brain-computer interface (BCI) • Braking intention • Feature combination • Electroencephalogram (EEG) • Driving

2.1 Introduction

Many kinds of driving assistant system have been developed for a driver's convenience and safety. Especially, the braking assistant technology is one of most important parts to assure a driver's safety. There were neurophysiological studies concerned with the use of brain signals for enhancing driving assistance systems. However, most of the studies focused on measuring and detecting drivers' physical conditions and mental states such as decreased concentration [1] or sleepiness [2–4] during a monotonous drive. Other studies focused on controlling the vehicle based on analysis of brain signals [5–11].

[1]**In a recent study, upcoming emergency situations during simulated driving were detected using event-related potentials (ERPs) [12]. This study demonstrated that neurophysiological correlates of emergency braking occur about 130 ms earlier than the corresponding behavioral responses related to the actual braking.** However, it is impossible to prove the feasibility of a driving assistant system based on a brain-computer interface in the real world, since the participants were exposed to a very reduced set of driving situations.

In this article, the feasibility of the driving assistant system is investigated based on the experiment with more diversified simulated driving situations than the previous study [12]. In our experiment, the participants were exposed to traffic situations including three kinds of emergency situations while they drove a virtual vehicle: the sudden stop of a leading vehicle, the sudden cutting-in of a vehicle from a neighboring lane, and the unexpected appearance of a pedestrian. In addition, there were driver's voluntary braking situations without any kinds of emergency stimulus. Electroencephalogram (EEG) signals were measured during the experiment and the neural responses of braking intention in various kinds of traffic situations were analyzed to investigate the differences of neural signals in emergency and non-emergency situations.

Three kinds of signal components were combined and used for the detection of driver's braking intention in this study. One signal component is the readiness potential (RP), a preparatory (i.e., pre-movement) component that indexes movement intention [13], ERPs (i.e., visual evoked potentials and P300 components), and event-related desynchronization (ERD) to distinguish diverse driving situations such as sharp braking.

The class-discriminability of univariate ERP features is investigated after the description of experimental paradigm and approach for signal analysis. After that,

[1]The parts have been taken verbatim from the author's prior publication [39] marked with bold in Introduction and Discussion Section.

we assess and compare the classification performances of the proposed feature combination method and the ERP feature based on the area under the specificity-sensitivity curve (AUC). The findings and experimental results are summarized with discussions at the end of this chapter.

2.2 Material and Methods[2]

2.2.1 Participants and Experimental Setup

Fifteen healthy individuals (all male and right-handed, age 27.1 ± 1.7 years) participated in this study. All participants had a valid driver's license and had driven for more than 3 years with no car-accident history. All had normal or corrected-to-normal vision. None of the participants had a previous history of psychiatric, neurological, or other diseases that might otherwise affect the experimental results. The experimental procedures were explained to each participant. Written informed consent was obtained from all participants before the experiment. The participants received a monetary reimbursement for their participation after the completion of the experiments. The participants were seated in a driving simulator cockpit (made by R.CRAFT in Korea) with fastened seat belts (the experimental apparatus is shown in Fig. 2.1).

Fig. 2.1 The experimental apparatus and environment

[2]The following section follows closely a prior published paper by the authors [39].

The virtual driving environment displayed on the screen was developed using the Unity 3D (Unity Technologies, USA) cross-platform game engine. This environment resembled an urban neighborhood without traffic lights and included autonomous (computer-controlled) vehicles as well as a vehicle to be steered by the participant. There was a six-lane road with many vehicles driving controlled by a simulator.

2.2.2 Experimental Paradigm

The participants were asked to drive a virtual vehicle using the accelerator and brake pedals and the steering wheel during one experimental session that lasted about 2 h. They were instructed to drive freely without getting into an accident, and to perform immediate braking to avoid crashes if necessary. The maximum speed of a vehicle was 120 km/h. We defined three kinds of braking situations. First, if an emergency situation occurred, the participants were instructed to press the brake pedal sharply. We defined this situation as sharp braking. Second, when the participant performed spontaneous braking to decrease a vehicle's speed, the vehicle decelerated **softly** (i.e., gradually). This situation was defined as soft braking. Finally, in many situations, the participants did not need to decrease the speed of their vehicle (i.e., normal driving). For all of the stimuli, the inter-stimulus interval (ISI) was between 4 and 18 s, and drawn randomly from a uniform distribution (see Fig. 2.2).

2.2.2.1 Sharp Braking Condition

There were three kinds of stimuli inducing sharp braking (emergency) situations. 1) The sharp braking by brake light condition: when the vehicle in front of the

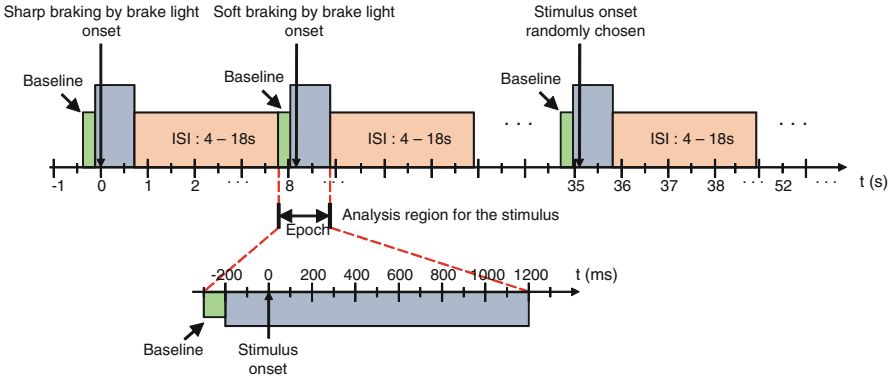

Fig. 2.2 Timing of the experimental paradigm

participant (lead vehicle) abruptly decelerated. A leading vehicle's brake light flashing was defined as the stimulus onset in this condition (see Fig. 2.3a). 2) The sharp braking by cutting-in condition: when a vehicle on the neighboring lane (side vehicle) abruptly cut in front of the participant's vehicle. The moment when the side vehicle came across the lane was defined as the stimulus onset (see Fig. 2.3b). 3) The sharp braking in the pedestrian condition: a pedestrian moved quickly toward the participants' vehicle from the side. The moment when the pedestrian left the sidewalk was defined as the stimulus onset (see Fig. 2.3c).

2.2.2.2 Soft Braking Condition

The soft braking condition was defined based on spontaneous braking in the absence of any stimulus. To slow down the vehicle, the participants spontaneously pressed the brake pedal. In this soft braking condition, the moment when the participant pressed the brake pedal was defined as the response onset (see Fig. 2.4a).

2.2.2.3 No-Braking Condition

The no-braking condition comprised three kinds of traffic situations. 1) One was normal driving: the participants just focused on driving, and no stimulus was given (see Fig. 2.4a). 2) The no braking by brake light condition: when a leading

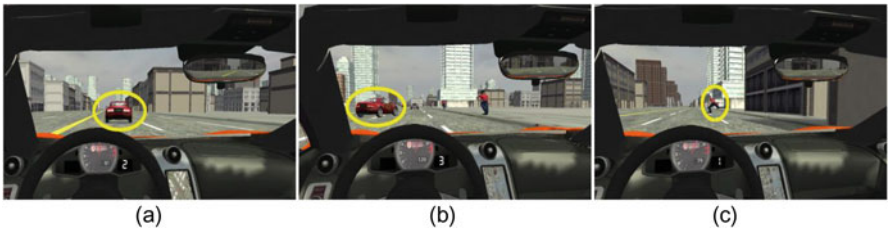

Fig. 2.3 Stimuli related to emergencies (**a**) Sharp braking by the leading vehicle's brake light, (**b**) Sharp braking by cutting-in, (**c**) Sharp braking by a pedestrian

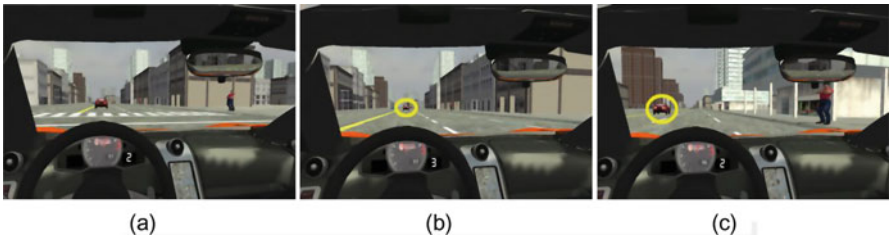

Fig. 2.4 Stimuli not related to emergencies (**a**) Soft braking or normal driving (no stimulus), (**b**) No braking by brake light, (**c**) No braking with brake light of neighboring vehicle

vehicle abruptly decelerated far away from a participant's car. In this condition, the participants did not have to press the brake pedal. The stimulus onset was defined as the moment in which the lead vehicle's brake light flashed (see Fig. 2.4b). 3) The no braking with brake light of neighboring vehicle condition: when the leading vehicle on a neighboring lane abruptly decelerated. In this condition, participants did not have to press the brake pedal. The side vehicle's brake light flashing was defined as the stimulus onset in this condition (see Fig. 2.4c).

2.2.3 Data Acquisition and Feature Extraction

The EEG signals were recorded using a multi-channel EEG acquisition system from 64 scalp sites based on the modified International 10–20 system [14]. We used Ag/AgCl sensors mounted on a cap (actiCAP, Brain Products, Germany). The ground and reference electrodes were located on scalp position AFz and the nose, respectively. The sampling rate was 1000 Hz throughout the experiments. The low and high cut-off frequencies were 0.1 and 250 Hz, respectively.

Electromyogram (EMG) signals were also acquired using a unipolar montage at the tibialis anterior muscle. The impedance of the EEG and EMG electrodes were kept below 10 kΩ. The EEG and EMG data were amplified and digitized using BrainAmp hardware (Brain Products, Germany).

Brake and gas pedal deflection markers were acquired at a 50 Hz sampling rate provided by the Unity 3D software. The time points of the braking response were defined based on the first noticeable brake pedal deflection that exceeded the jitter noise level.

The EEG signals were low-pass filtered at 45 Hz, and the EMG signals were band-pass filtered between 15 and 90 Hz. To remove line noise, we further applied a second-order digital notch filter at 60 Hz to the EMG signals.

The sampling rates of the physiological (EEG and EMG data) and mechanical (brake and gas pedal data) channels were down-sampled or up-sampled to 200 Hz for synchronization.

Segmentation was performed with 1500 ms length of epoch. Therefore, each epoch consists of 301 lengths of data points. After that, three different kinds of pre-processing and feature extraction methods were considered, corresponding to three different types of features (ERP, RP, and ERD) to be extracted from each epoch (see Fig. 2.5). The selection of time intervals (not necessarily of equal length) determined heuristically for each channel [15] (Ten-time intervals for ERP features and three-time intervals for RP features). These time intervals were selected only using training data. The same intervals were used for the test data. Each feature was computed for all electrodes (640 dimension of ERP features and 192 dimensions of RP features). Moreover, there are 6 ERD features for all channels. Thus, the combined feature vector has 838 dimensions (640 for ERP, 192 for RP, and 6 for ERD). The combination features were normalized by subtracting their empirical

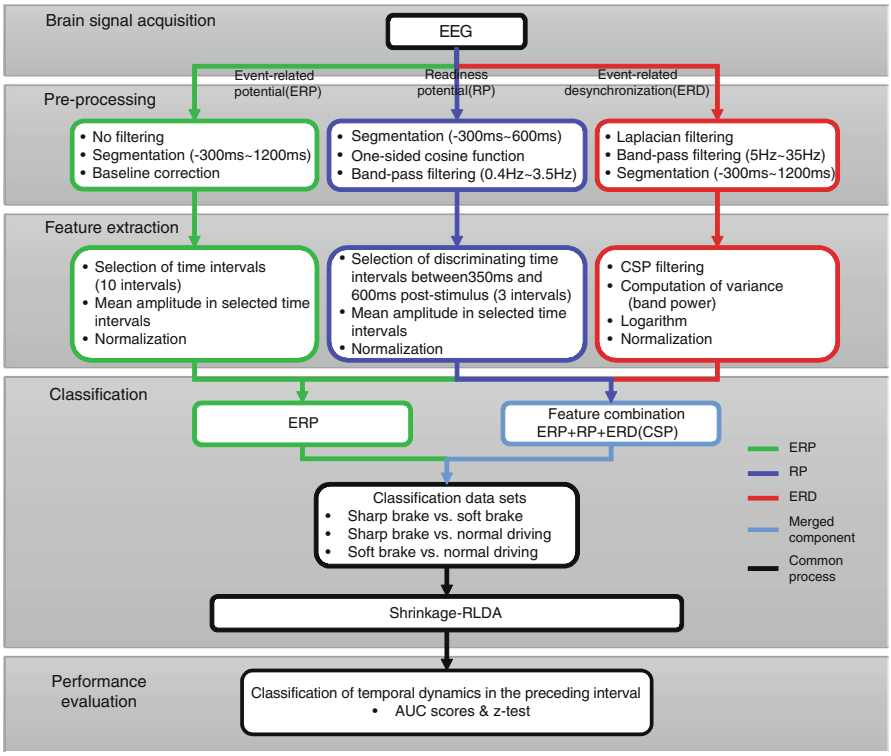

Fig. 2.5 Schematic flow of the data analysis process

means and dividing with their empirical standard deviations as estimated on the training sets to rescale the three kinds of features after concatenation of features. The test data sets were also normalized in the same way by subtracting the empirical means and dividing with the empirical standard deviations as estimated on the training data sets.

2.2.4 Event-Related Potentials and Area Under the Curve Analysis

The arithmetic mean of the extracted epochs of all 15 participants was computed to obtain grand-average ERP signals. Additionally, the discriminability of univariate (single-timepoint single-sensor) features with respect to the three predefined classes was investigated using the area under the curve (AUC) [16]. This analysis was conducted separately for each pair of classes. The AUC is symmetric around 0.5, where scores greater than 0.5 indicate that a feature has higher values in class 1 than in class 2 and scores smaller than 0.5 indicate

the opposite (smaller values in class 1) [16, 17]. The arithmetic mean of the AUC scores across the participants was calculated to obtain grand-average AUC scores.

2.2.5 *Classification*

As in [12], we evaluated the extent to which different feature modalities contribute to the overall decoding performance. Classifiers were trained on four kinds of modalities. These modalities were EEG (feature combination), EEG (only ERP features), EMG, and BrakePedal, which denotes the driver's actual brake pedal inputs. The first half of the epochs were used as the training set, and the second half were used as the test set. The entire analysis process including pre-processing is shown in Fig. 2.5.

The class discriminability of optimized combinations of spatio-temporal features was investigated using the regularized linear discriminant analysis (RLDA) classifier [18–21]. For regularization, the automatic shrinkage technique [22–25] was adopted. We had three classes of driving situations. For each pair of classes, we calculated the AUC scores of the RLDA outputs on the test set.

2.2.6 *Statistical Testing*

Whether a given AUC score was significantly different from 0.5 (that is, chance level) on the population level was assessed by means of two-sided Wilcoxon signed rank test [26]. To assess whether two scores AUC1 and AUC2 were significantly different from each other, the difference AUC1-AUC2 was also tested for being nonzero using the same two-sided Wilcoxon signed rank test. Bonferroni-correction was implemented to obtain reliable p-values [27]. The correction factor was 301 (time instants) $\times$ 64 (electrodes) = 19,264 in the ERP analysis. For other channels, we used a correction factor of 301 (time instants). Features whose p-value was smaller than 0.05 are considered statistically significant.

2.3 Results[3]

2.3.1 *ERP Analysis Related to Braking Conditions*

Each of the three classes of driving situations induced a specific cascade of brain activities representing low- and high-level processing of the (visual) stimulus as

[3]The following section follows closely a prior published paper by the authors [39].

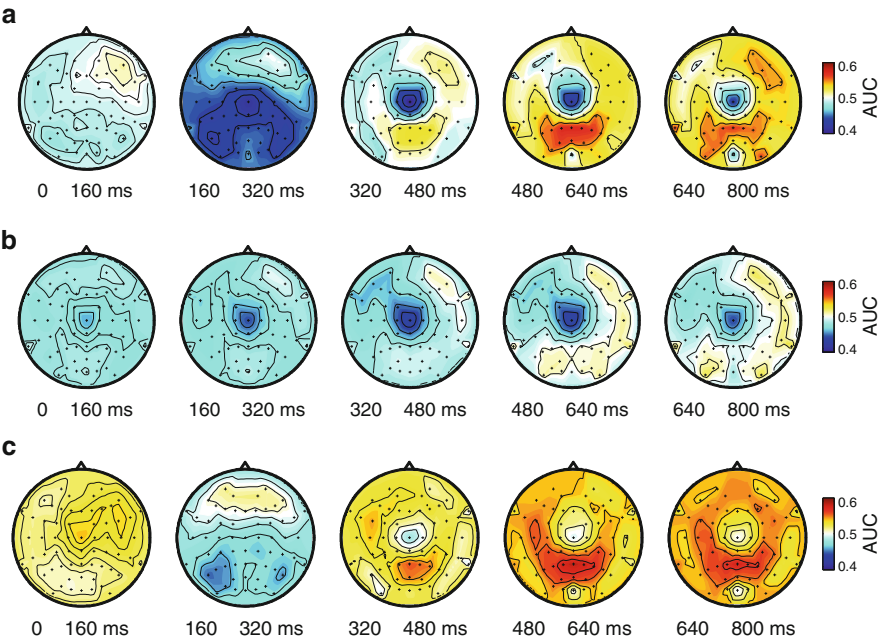

Fig. 2.6 Topographical maps of grand-average AUC scores calculated from average ERPs in five temporal intervals. (**a**) Sharp braking vs. no braking, (**b**) Soft braking vs. no braking, (**c**) Sharp braking vs. soft braking

well as motor preparation and execution. The same number of braking situations was induced for all braking types and subjects. However, after filtering and artifact rejection, the number of trials used in the analysis of neural correlates varied. On average, 57.1 ± 12.3 trials related to sudden stops of leading vehicles, 60.5 ± 8.1 trials related to cutting-in of leading vehicles, and 44.7 ± 5.6 trials related to sudden appearance of pedestrians were used in the data analysis. Half of the trials were used for training the classifier, and the remaining trials were used to evaluate the decoding performance.

We investigate the spatio-temporal ERP sequences reflecting the class-discriminative brain processes. Figure 2.6 shows topographical maps of grand average AUC scores in five subsequent 160 ms long time intervals. Figure 2.6a shows the AUC scores related to the difference between sharp braking and no braking. The feature value of sharp braking is higher than that of no braking ($AUC > 0.5$) in the time interval between 320 and 800 ms post-stimulus in parietal areas. The AUC score is maximal in the time interval between 480 and 640 ms ($z > 6.8$, $p \approx 0$). The electrode having the highest AUC score (0.58) between 480 and 640 ms post-stimulus is Pz. On the other hand, the higher feature value of no braking than that of sharp braking ($AUC < 0.5$) is observed in the time interval between 160 and 320 ms in lateralized occipital areas ($z < -6.8$, $p \approx 0$), and in

the time interval between 320 and 800 ms in central areas ($z < -6.8$, $p \approx 0$). The electrode Cz shows the lowest AUC score (0.40) between 320 and 480 ms post-stimulus.

Figure 2.6b shows the AUC scores related to the difference between soft braking and no braking. The higher feature value of no braking than that of soft braking is observed over the entire time interval in central areas. The lowest AUC score (0.43) is observed in the electrode Cz between 320 and 480 ms post-stimulus. The electrode TP10 has the highest AUC score (0.53) between 640 and 800 ms post-stimulus. Figure 2.6c depicts the AUC scores related to the difference between sharp braking and soft braking. Here, the highest feature value of sharp braking is observed in parietal/occipital areas in the time interval between 480 and 640 ms ($z > 6.8$, $p \approx 0$), while a highest feature value of soft braking around electrode Cz is observed in the time interval between 320 and 640 ms ($z < -6.4$, $p \approx 0$). The electrode P7 has the lowest AUC score (0.44) between 160 and 320 ms post-stimulus while the highest AUC score (0.59) is observed in electrode Pz between 480 and 640 ms post-stimulus. Thus, sharp braking elicits stronger feature values than soft braking and is moreover characterized by the additional presence of visual-evoked potentials and a P300 component. Note that it is impossible to distinguish emergency braking situations from normal driving (e.g., no-braking events) before the stimulus. Therefore, the classification before stimulus onset must be at chance level as indicated by AUC scores of 0.5 (see Fig. 2.6).

2.3.2 Comparison of Classification Results Based on ERP Features and a Novel Feature Combination

The results of the classification analyzes using multivariate features are shown in Fig. 2.7. These features were extracted from stimulus-locked segments. The classification performance was measured in terms of AUC scores achieved by the outputs of RLDA classifiers on test data. The classifiers were trained to distinguish two of the three classes. Thus, there were three different class combinations to consider.

Figure 2.7 provides a time-resolved assessment of the classification performance, where the AUC score at each time point represents the accuracy achievable using the preceding 1500 ms long segment of data. The boxplots in Fig. 2.7 shows the distribution of reaction times (defined as the first above-threshold brake pedal deflection) for the sharp braking condition. Note that for the soft braking condition artificial stimulus onsets were sampled from the same distribution.

The classification results with respect to distinguishing sharp braking and no braking based on ERP features are similar to the results previously achieved by [12] (see Fig. 2.7a). The AUC scores of both single features and the proposed feature combination based on EEG exceed 0.6 after 260 ms post-stimulus. In addition, the score of both EMG and brake signal exceed 0.6 after 320 and 580 ms post-stimulus,

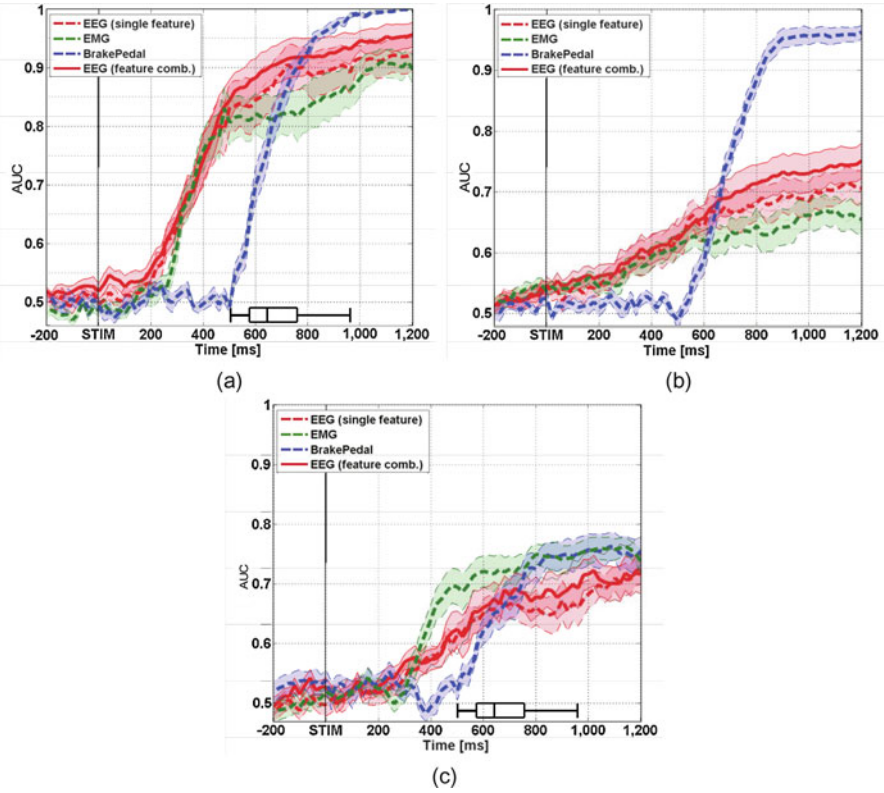

Fig. 2.7 Classification performances based on ERP feature and novel feature combination. The distribution of reaction times for the sharp braking condition is depicted by boxplots. The areas of *shaded color* represent standard errors of the mean (SEM) of the AUC scores

respectively. The AUC of the EEG using the novel feature combination exceeds a value of 0.9 by 80 ms ahead of using ERP features alone.

The novel feature combination shows significantly better performance compared to the ERP features from 640 to 1200 ms post-stimulus ($p < 0.05$, the largest difference at 960 ms, $z_{min} = 5.4$, $p \approx 10^{-2}$).

The performance with respect to distinguishing the soft braking and no braking conditions is presented in Fig. 2.7b. For EEG, the scores using the novel feature combination are considerably higher than the scores obtained from using ERP features only. For ERP only features, the AUC exceed 0.6 at 440 ms post-stimulus, while for the novel feature combination it is 340 ms post-stimulus. In addition, the novel feature combination achieves significantly better performance than ERP-only features from 760 to 1200 ms post-stimulus ($p < 0.05$, the largest difference at 840 ms, $z_{min} = 5.8$, $p \approx 10^{-3}$).

Finally, the performance in classifying sharp braking and soft braking based on ERP features is dramatically lower compared to using EMG features in the entire

time interval considered. On the other hand, the performance is improved by means of the proposed combination of EEG features, although the achieved scores are still lower than those obtained from EMG. The novel feature combination achieves an AUC score of 0.6 by 120 ms earlier than the corresponding ERP-only features. It is significantly better than the ERP features alone in the interval from 380 to 440 ms post-stimulus (significant with $p < 0.05$, the largest difference at 420 ms, $z_{min} = 5.6$, $p \approx 10^{-3}$). These results are presented in Fig. 2.7c. These results confirm that our novel feature combination is more informative for the detection of drivers' braking intentions than ERP features alone.

2.4 Discussion and Conclusion

A positive signal similar to the typical P300 component is observed in a broad parieto-occipital region for all kinds of visual stimulus types in the analysis results of spatio-temporal ERP pattern. This neurophysiological properties of the ERPs are caused by our experimental paradigm similar to the classical oddball paradigm [28] in a respect that the emergency stimuli (oddball) are presented randomly during normal driving (normal ball).

The neurophysiological responses to the two kinds of braking (soft and sharp) show the different amplitude of positivity in the parieto-occipital region and this difference was helpful to distinguish these two kinds of situations.

There are three kinds of stimuli in case of sharp braking condition and the ERP patterns across the scalp evoked by these stimuli were different. However, we considered these patterns as the same class because these patterns evoked by pseudo-oddball paradigm and we could observe a positive potential in the parieto-occipital region [29] for all kinds of stimuli. In the end, we were able to detect the driver's braking intention regardless of what kind of traffic situations occurred.

A negative signal in a central region (especially at the Cz electrode) evoked by the planning processes in the motor system related to the act for pushing the gas or brake pedal (i.e., the readiness potential) [30, 31] in all braking situations. Different negative deflection was evoked by reactive and spontaneous movements at the foot area of the motor cortex during and prior to muscle movement [6, 32, 33]. In addition, the reactive movement and spontaneous movement had a different start time of the negative deflection in the central region. The reactive movement was similar to cue-based motor execution and corresponding to sharp braking. On the other hand, the spontaneous movement was similar to self-paced motor execution and corresponding to soft braking. **The start point of the pre-motion negative deflection in the central motor area related to a reactive movement was later than the start point of the spontaneous movement, in line with** [13, 34, 35]**. Thus, as aforementioned, we used the readiness potential as a feature because it provided important movement-related information.**

There is about 150 ms of the time difference between ERD starts and EMG onset (i.e., ERD is faster than EMG) [36, 37]. **The ERD was observed prior to the depression of the brake pedal (i.e., right foot movement from the gas to the brake pedal). Interestingly, the self-initiated movement (i.e., soft braking) and the movement triggered by an external stimulus (i.e., sharp braking) differed in aspects (i.e., magnitude and slope) of the band power. Thus, the ERD information related to foot movement for braking was also used as a feature.**

The prediction performance of the sharp braking and soft braking conditions are assessed in time series based on AUC score. The AUC score of EEG increased faster than that of EMG. Although the peak AUC score of EMG is lower than that observed in a previous study [12], the trends of results are similar to a previous study.

The feasibility of smart vehicle technology for detection of driver's braking intention was verified in this study by detecting the participants' braking intention robustly in all of the simulated traffic situations. Especially, the prediction performance based on the novel feature combination is better than the prediction performance based on ERP features alone reported in a previous study [12] although the classification performance for sharp braking and soft braking based on EEG features was lower than the performance based on the EMG. Some important environmental stimuli (e.g., vehicle vibration, auditory stimuli) were omitted in our setting. However, it is complemented by the work of [3, 38], which shows that results identical or better to those of [12] can be obtained in a real-world setting.

Together, our study and [38] provide a converging evidence suggesting that smart vehicles that have an automatic braking assistance system integrating neurophysiological responses could be developed in practice. The methods and results sections have been taken from an own prior publication after summarization. For further details on the experiment and analysis, see [39].

Acknowledgements This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2015R1A2A1A05001867). The authors acknowledge the use of text from the own prior publication [39] in this article. Jeong-Woo Kim and Seong-Whan Lee thank their co-authors for allowing them to use materials from prior joint publication [39].

References

1. Schmidt EA, Schrauf M, Simon M, Fritzsche M, Buchner A, Kincses WE (2009) Driver's misjudgement of vigilance state during prolonged monotonous daytime driving. Accid Anal Prev 41(5):1087–1093
2. Papadelis C, Chen Z, Kourtidou-Papadeli C, Bamidis PD, Chouvarda I, Bekiaris E, Maglaveras N (2007) Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. Clin Neurophysiol 118(9):1906–1922
3. Kohlmorgen J, Dornhege G, Braun M, Blankertz B, Müller KR, Curio G, Hagemann K, Bruns A, Schrauf M, Kincses W (2007) Improving human performance in a real operating environment through real-time mental workload detection. In: Toward brain-computer interfacing. MIT Press, Cambridge, MA, pp 409–422

4. Schmidt WE, Kincses WE, Schrauf M, Haufe S, Schubert R, Curio G (2007) Assessing driver's vigilance state during monotonous driving. In: Proceedings of the 4th International symposium on human factors in driving assessment, training, and vehicle design, Washington, USA, pp 138–145
5. Hood D, Joseph D, Rakotonirainy A, Sridhara S, Fookes C (2012) Use of brain computer interface to drive: preliminary results. In: Proceedings of the 4th International conference on automotive user interfaces and interactive vehicular applications 2012, Eindhoven, Netherlands, pp 103–106
6. Khaliliardali Z, Chavarriaga R, Gheorghe LA, Millán JdR (2012) Detection of anticipatory brain potentials during car driving. In: Proceedings of the IEEE conference on EMBS 2012, San Diego, USA, pp 3829–3832
7. Luzheng B, Nini L, Xin-an F (2012) A brain-computer interface in the context of a head up display system. In: Proceedings of the ICME International conference on complex medical engineering 2012, Kobe, Japan, pp 241–244
8. Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 167(1):82–90
9. Renold H, Chavarriaga R, Gheorghe LA, Millán JdR (2014) EEG correlates of active visual search during simulated driving: an exploratory study. In: Proceedings of the IEEE International conference on systems, man and cybernetics, San Diego, CA, USA, 5–8 Oct 2014, pp 2815–2820
10. Zhang H, Chavarriaga R, Gheorghe LA, Millán JdR (2013) Inferring driver's turning direction through detection of error related brain activity. In: 35th Annual international conference of the IEEE engineering in medicine and biology society, Osaka, Japan, 3–7 July 2013, pp 2196–2199
11. Gheorghe LA, Chavarriaga R, Millán JdR (2013) Steering timing prediction in a driving simulator task. In: 35th Annual international conference of the IEEE engineering in medicine and biology society, Osaka, Japan, 3–7 July 2013, pp 6913–6916
12. Haufe S, Treder MS, Gugler MF, Sagebaum M, Curio G, Blankertz B (2011) EEG potentials predict upcoming emergency brakings during simulated driving. J Neural Eng 8(5):056001
13. Shibasaki H, Hallett M (2006) What is the bereitschaftspotential? Clin Neurophysiol 117(11):2341–2356
14. Sharbrough F, Chartrian GE, Lesser RP, Lüders H, Nuwer M, Picton TW (1991) American electroencephalographic society guidelines for standard electrode position nomenclature. J Clin Neurophysiol 8(2):200–202
15. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 25(1):41–56
16. Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874
17. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
18. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Hum Genet 7(2):179–188
19. Duda RO, Hard PE, Stork DG (2000) Pattern classification. Wiley, New York
20. Lemm S, Blankertz B, Dickhaus T, Müller KR (2011) Introduction to machine learning for brain imaging. NeuroImage 56(2):387–399
21. Tomioka R, Müller KR (2010) A regularized discriminative framework for EEG analysis with application to brain-computer interface. NeuroImage 49(1):415–432
22. Friedman JH (1989) Regularized discriminant analysis. J Am Stat Assoc 84(405):165–175
23. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR (2011) Single-trial analysis and classification of ERP components – a tutorial. NeuroImage 56(2):814–825
24. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal 88(2):365–411
25. Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4(1):32

26. Gibbons JD, Chakraborti S (2011) Nonparametric statistical inference. Marcel Dekker, New York
27. Bonferroni CE (1936) Teoria Statistica delle Classi e Calcolo delle Probabilitá. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3–62
28. Sutton S, Braren M, Zubin J, John ER (1965) Evoked-potential correlates of stimulus uncertainty. Science 150(3700):1187–1188
29. Kim I-H, Kim J-W, Haufe S, Lee S-W (2013) Detection of multi-class emergency situations during simulated driving from ERP. In: 2013 IEEE International winter workshop on brain-computer interface, Jeongseon, Korea, pp 49–51
30. Kornhuber HH, Deecke L (1965) Hirnpotentialänderungen bei Willkürbeweguugen und passive Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. Pügers Arch 284:1–17
31. Penfiled W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain J Neurol 60:389–443
32. Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. Brain J Neurol 106(3):623–642
33. Haggard P, Eimer M (1999) On the relation between brain potentials and the awareness of voluntary movements. Exp Brain Res 126(1):128–133
34. Dornhege G, Blankertz B, Curio G, Müller KR (2004) Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Trans Bio-Med Eng 51(6):993–1002
35. Blankertz B, Dornhege G, Schäfer C, Krepki R, Kohlmorgen J, Müller KR, Kunzmann V, Losch F, Curio G (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans Neural Syst Rehabil Eng 11(2):127–131
36. Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110(11):1842–1857
37. Leocani L, Toro C, Zhuang P, Gerloff C, Hallett M (2001) Event-related desynchronization in reaction time paradigms: a comparison with event-related potentials and corticospinal excitability. Clin Neurophysiol 112(5):923–930
38. Haufe S, Kim J-W, Kim I-H, Sonnleitner A, Schrauf M, Curio G, Blankertz B (2014) Electrophysiology-based detection of emergency braking intention in real-world driving. J Neural Eng 11(5):056011
39. Kim I-H, Kim J-W, Haufe S, Lee S-W (2015) Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. J Neural Eng 12(1):016001

Chapter 3
Benefits and Limits of Multimodal Neuroimaging for Brain Computer Interfaces

Siamac Fazli, Min-Ho Lee, Seul-Ki Yeom, John Williamson, Isabella Schlattner, Yiyu Chen, and Seong-Whan Lee

Abstract Recently there has been a surge of interest for combining data from various sources in neuroscience, and also in other scientific domains. In this article we examine some of the benefits as well as limitations that arrise, when various neuroimaging techniques are employed for Brain-Computer Interfacing. In particular we review how setup costs can be reduced for multimodal systems, the NIRS response delay minimized and furthermore show that NIRS can help in the robust detection of the idle state.

Keywords BCI • EEG • NIRS • Multi-modal neuroimaging • Sensor reduction • Idle-state detection

3.1 Introduction

Brain-computer interfaces (BCIs) are systems that enable humans to communicate with machines by means of their thoughts. The goal of BCI is to offer people with motor disabilities such as tetraplegia or even locked-in syndrome a communication tool to the external world by bypassing the brain's normal output pathways, such as peripheral nerves and muscles [61].

One possible input that can be employed by BCI is neural activity recorded from the scalp. In 1924 Hans Berger recorded the first human Electroencephalogram [6]. Electroencephalography (EEG) is a non-invasive method that records electrical signals generated by the brain through electrodes, which are placed at different positions of the scalp. Some of the advantages of EEG are that it is non-invasive, has good temporal resolution, the equipment is comparatively cheap, and it is transportable. The disadvantages of EEG are its limited spatial resolution as well as low signal-noise-ratio.

S. Fazli (✉) • M.-H. Lee • S.-K. Yeom • J. Williamson • I. Schlattner • Y. Chen • S.-W. Lee (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: fazli@image.korea.ac.kr; swlee@image.korea.ac.kr

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_3

Near-infrared spectroscopy (NIRS) is a non-invasive optical technique that measures the absorbance of light in tissue at different wavelengths in the spectral region from 700 to 1,000 nm [32]. Accordingly, it allows the determination of changes in the concentration of oxygenated hemoglobin ([HbO]), deoxygenated hemoglobin ([HbR]) and blood volume due to neural activity [59]. NIRS has a good spatial resolution when compared to EEG. However, due to the latency of brain hemodynamics its temporal resolution is very limited and similar to that of functional Magnetic Resonance Imaging (fMRI) [7, 30].

Lately, NIRS is receiving a lot of attention. While earlier NIRS systems were only able measure very few channels, today whole-head systems become more available. The NIRS technology has been remarkably developed, newer NIRS technologies have become a promising tool for neuroscientific studies. Consequently, the use of NIRS in neuroscience has been constantly increasing and well developed in several subfields. In functional imaging studies, some researchers have even developed a wearable multi-channel fNIRS system [45] which can be used in unrestrained settings.

In 2004, Coyle and colleagues first demonstrated the feasibility of a NIRS-based optical BCI [9]. Since then a number of groups have followed the idea of using optical data for BCI systems [10, 33, 54, 62].

3.1.1 Multimodal Neuroimaging

Interest in non-invasive multimodal imaging in neuroscience has been increasing in the past 20 years since Ives and his group of researchers were the first to record EEG signals simultaneously with an fMRI scan [31]. The combination of fMRI's hemodynamic spatial resolution with the temporal resolution of EEG is an attractive option for many researchers and clinicians. Besides just imaging with EEG-fMRI data, other multimodal imaging setups have been developed and applied in clinical and research settings using combinations of other modalities which include MEG, NIRS, PET, CT and MRI [3, 4, 14, 44, 51].

The basic assumption behind multimodal acquisition of data is that one modality can compensate or complement the data for the other modality's deficiencies. In the case of EEG and fMRI, the complementary relationship of their respective temporal and spatial resolution clearly justifies their multimodal combination. Besides combinations that complement the canonical shortcomings for each modality, compensatory combinations can also provide imaging solutions for researchers. For example, simultaneous recordings of EEG and MEG data have proven useful in improving source estimation for these spatially poor modalities by compensating for each other's specific deficiencies in source localization [4]. Similarly, fNIRS and fMRI have been used simultaneously to better understand the hemodynamic response by combining fNIRS's greater biochemical specificity and sensitivity to the microvasculature with fMRI's more robust spatial resolution [51]. Additionally, most modern PET scanners are combined with CT scans to acquire better anatomical

imaging, but since the image acquisition in these modalities are performed sequentially, not simultaneously, and because both require radiation exposure, development of an fMRI and PET multimodal technique is in development to increase safety by reducing the subject's exposure to radiation [44].

As Friston points out in his review of multimodal imaging [24], despite the excitement surrounding multimodal imaging in neuroscience, its application is still not commonplace. Certainly, cost is an issue as well as the fact that many clinical and research questions can still be adequately investigated with a unimodal method. Another reason for the limited application of multimodal imaging is because of the many challenges in the analysis of simultaneously obtained data sets which measure different aspects of an imprecisely known source. A review of analysis techniques and data integration is beyond the scope of this chapter but for a good starting point of review, see [7, 12, 13, 18, 43].

Despite the challenges, complications, and increased cost, multimodal imaging is blossoming within specific disciplines as a powerful research tool by allowing researchers and clinicians to reevaluate their limitations of investigation and providing opportunity for deeper understanding. For example, simultaneous EEG-fNIRS imaging is proving to be a more versatile, lower-cost option in many fields like psycholinguistics [60], optical topography [16], and certainly in BCI which will be the focus for the rest of this chapter.

3.1.2 Combining EEG and NIRS for BCI

In a recent study, simultaneous measurements of NIRS and EEG were recorded during a real-time sensorimotor rhythm (SMR)-based BCI feedback experiment [23]. The feedback was EEG-based, but in further offline simulations the classification accuracies for each signal domain were both evaluated separately as well as in combination. This combination was performed by means of meta-classifiers. The meta-classifiers weigh the linear classifiers of the individual measurements according to training data samples [23]. The unimodal analysis showed that single-trial EEG classification accuracy was superior, when compared to the NIRS chromophores. However, the multimodal analysis revealed that the usage of meta-classifiers, combining features from EEG and NIRS, improved the classification accuracy significantly. This increase was observed in over 90 % of the considered subjects and led to a significant average performance increase of 5 % across subjects.

To further examine the degree of independence between the NIRS and EEG-based classifier outputs, their outputs can be restricted to values between 0 and 1 and their mutual information $\mathcal{I}$ can be estimated. Mutual information is an information theoretic measure, which estimates the information that two random variables share. It can be expressed in terms of conditional entropies of random variables X and Y:

$$\mathcal{I}(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{3.1}$$

The conditional entropy $H(X|Y)$ quantifies the remaining entropy of X, after the value of Y is known. If $H(X|Y) = H(X)$, then $\mathcal{I}(X;Y) = 0$: the variables are independent. On the other hand, if X and Y are identical, then $H(X|Y) = 0$ and hence $\mathcal{I}(X;Y) = H(X)$. $\mathcal{I}(X;Y)$ is symmetric and its values are in the range of 0 and 1: $\mathcal{I}(X;Y) = \mathcal{I}(Y;X) \in [0;1]$ [41].

With increasing accuracy of NIRS chromophores, the mutual information also rises. Similarly, with increasing EEG accuracy, the mutual information rises as well. The intuition is clear: with increasing accuracy, both classifier outputs will correctly predict and thus share the majority of class labels. However, the normalized mutual information did not reach values above 0.4 bit for any subject. To investigate, whether EEG and NIRS classifiers misclassify the same trials, the EEG classification accuracy of all trials was examined in relation to the NIRS classification accuracy of trials, where [HbO] was correct/incorrect. The results indicated, that EEG classification is largely invariant to the classification performance of [HbO] and [HbR]. In other words, *EEG and NIRS mostly misclassify different trials*. It can therefore be deduced that the individual methods in fact complement each other in terms of *information content* [23].

3.1.3 Current Issues of Multimodal BCIs

NIRS measures changes in local concentration of oxygenated and deoxygenated hemoglobin ([HbO] and [HbR]) in the cerebral blood. Previous studies reported that the hemodynamic response typically peaks about 6s after stimulus onset [48, 58] and then dips back down over the course of several more seconds as homeostasis is re-established. Therefore, NIRS-based BCI systems require a paradigm with relatively long Stimulus Time Intervals (STIs) and Inter Stimulus Intervals (ISIs), when compared to EEG experiments [29, 42]. This inherent latency, caused by the slow hemodynamic response presents an important challenge for multimodal EEG+NIRS-based BCI systems today: it would be prohibitively inconvenient to wait 6s following a mental command for a BCI to respond. Most of the previous NIRS-based BCI studies used time-averaged hemodynamic concentration features [9, 54], which inevitably leads to low Information Transfer Rates (ITRs). X. Cui et al. proposed a multivariate pattern classification technique with different feature spaces (signal history gradient and spatial pattern) to reduce this delay [11]. While it reduces the latency of decoding a behavioral state from 6s to 4s, it still requires extensive offline processing. An EEG-NIRS multimodal system is expected to enhance BCI accuracy. At the same time a loss of ITR should be avoided, if possible. Therefore, future studies investigating the trade-off relationship between accuracy and ITR will be of benefit for the usability of multimodal BCI systems.

EEG-NIRS experiments require a long setup time. Setup time includes the setup of hardware, such as applying conducting gel to EEG electrodes and setting up NIRS optodes on the scalp, but also the estimation of subject-dependent filters

for decoding. Dry electrodes [28, 46] as well as EEG-based subject independent decoding [19–21, 36] already adress this issue for the domain of EEG. However, for multimodal experiments, the preparation time is significantly extended, when compared to a unimodal experiment. This long setup time often leads to subject fatigue, which can result in subjects failing to concentrate on their task.

While the EEG signal can be influenced by noise artifacts such as electro-oculogram (EOG) and electromyogram (EMG), NIRS signal quality is strongly influenced by the density of hair which can cause poor optical contact [34]. Generally EEG-based BCI systems employ scalp areas, that are covered in hair, such as the motor-cortical areas for SMR-based BCI, occipital areas for steady-state visual evoked potential (SSVEP)-based BCIs, among others. In contrast, a number of proposed NIRS-BCI systems were based on prefrontal areas which are free of hair [5, 35, 52]. In a recent multimodal EEG-NIRS study, the NIRS electrodes were positioned over the prefrontal region with the EEG electrodes over the motor cortex regions [35]. In contrast Fazli et al. have set the EEG electrodes and NIRS optodes together within the same region [23].

In summary, for simultaneous data recording the optimal channel configuration should be predetermined, depending on the task at hand. Whether or not EEG and NIRS should monitor the same region [23] or are independently located [35] needs to be decided on a case by case basis.

The majority of sensory motor rhythm (SMR)-based BCIs require the subject to react to a cue, a so-called *cue-paced* or *synchronous BCI*. These synchronous BCI systems are disadvantageous for a broad range of applications outside the laboratory and it is therefore important to design systems that are able to detect *asynchronous* (or *self-paced*) actions. In asynchronous BCI paradigms the users' intentions are divided into *active* and *idle states*. During the *active state* the user concentrates on the specific mental tasks to operate the system, while during the idle state no interaction is required. Clearly, the robust detection of idle states is highly important for the successful operation of such asynchronous BCIs. In Sect. 3.2.3 we show that the use of multimodal neuroimaging can successfully assist in this task.

The above mentioned challenges, namely the reduction of sensors in multimodal setups, the minimization of the NIRS response as well as the robust idle state detection will be examined in the following by revisiting a previously recorded dataset [23].

3.2 Techniques for Solving Previous Issues

3.2.1 Sensor Reduction for Multimodal Setups

To produce high accuracy decoding most BCI research is based on whole-head multi-channel as well as multivariate analysis techniques. However, large numbers of electrodes require a longer time spent in channel preparation. Portable systems

with fewer channels have become essential when applying BCIs to everyday life and home applications. As mentioned above, combined EEG and NIRS measurements increase setup times significantly. For EEG-based BCIs common spatial patterns (CSP) is a widely used method for effective feature extraction [25, 47]. Furthermore, various EEG-based channel selection (CS) algorithms such as Recursive Channel Elimination (RCE) [50], single-channel [53], pre-defined CS [49] have been developed. Although those channel selection algorithms have been widely used in EEG-based BCIs, they have so far not sufficiently been investigated for NIRS-based BCIs.

Here, we provide a review on a recently proposed statistical channel selection heuristic, which efficiently reduces the computation time without loss of performance for high dimensional NIRS data [38].

The NIRS-System (NIRScout 32–32, NIRx) was equipped with 32 sources and 32 detectors convolving to 108 measurement channels which covered a whole-head area. The optical (sources and detectors) probes were located by the international 10–20 system and data was collected with a sampling frequency of 6.25 Hz. Five subjects were instructed to perform cued left/right hand finger movements. The stimulus interval was set to 20 ± 1.5 s, and the order of cues was randomized. Each subjects' dataset consists of a single sessions with 120 trials of hand gripping.

Low-frequency artifacts such as respiration, heart rate and Mayer waves were filtered using a low-pass, 3rd order Butterworth filter at 0.2 Hz. The changes of oxygenated and deoxygenated hemoglobin were calculated, based on the modified Beer-Lambert law [37]. The baseline correction was performed (−2s to 0), and mean amplitude features were calculated. The point biserial correlation coefficient was applied to individual channels to find the differences between classes [40, 56]:

$$r(x) = \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \cdot \frac{\text{mean}\{x_i | y_i = 1\} - \text{mean}\{x_i | y_i = 2\}}{\text{std}\{x_i\}} \tag{3.2}$$

where y_i represents the class label and N the total number of datapoints in target and non-target classes. The signed-r^2-values are defined as $\text{sign}(x) \cdot r(x)^2$. Sgn r^2 values were calculated for each channels and time point. Not only the maximum value of sgn r^2, but also the area under the absolute r^2 curve were computed for the entire time series of X to devise a contribution score for each channel. The contribution scores as well as the optimal thresholds of keeping channels were estimated by 6-fold nested cross-validation.

Following this methodology, only a few number of informative channels were selected (≈5–6 channels per subject) from the 108 dimensional NIRS data, while the classification accuracy was maintained. This result suggests that applying CS algorithms to high dimensional NIRS data can lead to significantly reduced preparation time of NIRS-EEG experiments.

3.2.2 Reduction of the NIRS Response Time

The temporal delay inherent to cerebral hemodynamics has so far hindered the simultaneous use of NIRS and EEG in real-time applications. While the multimodal combination of EEG and NIRS has previously been realized in the domain of BCI and yielded beneficial results, only trials with relatively long inter-stimulus intervals (ISI) were considered. Here we examine whether NIRS data can also help to increase classification performance for the case of fast-paced paradigms with shorter ISIs.

Data of a previously published study was employed [23], where fourteen healthy, right-handed volunteers (aged 20–30) participated. Each subject was instructed to perform left or right hand motor imagery, depending on the direction of the arrow on the monitor. The experiment was performed with an ISI of 14.5 ± 1.5 s during the calibration phase and an ISI of 7 ± 0.5 s during the fast feedback phase. During both the calibration phase as well as the fast feedback phase real-time EEG-based feedback was provided.

To examine how well the EEG and NIRS data classifies the given tasks, time courses were analyzed with the help of a moving window (width 1 s, step size = 500 ms) that was applied to the calibration data with long ISIs as well as the data from the *fast feedback phase*. Time courses of [HbO] and [HbR] were averaged within the 1 s time window and used as features to train a linear discriminant analysis classifier with shrinkage (RLDAshrink). Validation was performed by cross-validation with an 8-fold chronological split for each window separately. Furthermore, for combining EEG and NIRS chromophores various LDA meta classifiers were estimated using the outputs from the cross-validated, respective signals. A meta classifier combining EEG with [HbO], another combining EEG with [HbR] and a third combining EEG with both chromophores were computed.

The classification results of the best time window for the long and short ISI data can be obtained from Table 3.1. As expected, EEG data is more discriminable than NIRS chromophores. Furthermore, we show that classification accuracies of [HbO] for short ISIs are on par with long ISI data, thus confirming previous findings [63]. To investigate this interesting finding further, the classifier outputs for the fast feedback phase were averaged across trials and subjects. The results can be seen in Fig. 3.1, which presents the grand average time-course of task performance for EEG and the two NIRS chromophores ([HbO] and [HbR]). The light green line indicates the time point at which the means of the classification outputs deviate significantly ($p < 0.001$). Two-sample t-tests with the null hypothesis of equal means were performed for each timepoint and Bonferroni corrected. While the EEG shows an earlier timepoint than the two NIRS chromophores (EEG = 860 ms, HbO = 1,240 ms, HbR = 1,560 ms), the NIRS chromophores show astonishingly early deviations. The results indicate that the latency of the NIRS chromophores is short enough to assist EEG-based classification, even within a real-time BCI feedback environment: NIRS may assist and enhance BCI feedback performance significantly, even shortly after stimulus onset.

Table 3.1 Uni-/multimodal classification results with long-ISIs as well as short ISIs (i.e. fast-paced feedback) of all individual subjects for EEG, [HbO] and [HbR]. Stars indicate the significance of improvement, based on a paired t-test with the hypothesis of equal means (** stands for $p < 0.05$). EEG+O, EEG+R and EEG+O+R stand for the combinations of EEG and [HbO], EEG and [HbR] as well as EEG, [HbO] and [HbR]

	Unimodal-long ISI			Unimodal-short ISI			Multimodal – short ISI		
Subj	EEG	[HbO]	[HbR]	EEG	[HbO]	[HbR]	EEG+O	EEG+R	EEG+O+R
a	55.0	77.5	61.0	60.5	75.8	53.0	76.7	59.5	78.7
b	99.0	63.0	58.0	99.5	49.2	48.6	99.5	99.5	99.5
c	68.0	58.0	59.5	65.5	77.7	47.0	72.1	64.5	72.2
d	75.0	62.0	61.0	75.3	63.0	43.2	78.2	74.7	77.6
e	83.5	78.5	61.5	88.7	92.4	90.3	96.1	91.5	96.6
f	59.5	57.5	56.0	57.7	51.6	51.0	58.1	58.1	57.8
g	88.5	92.5	89.5	72.8	90.7	93.0	93.1	92.2	93.6
h	94.5	74.5	68.5	92.1	84.5	72.9	92.2	92.3	92.3
i	80.0	56.0	61.0	97.1	47.7	50.8	97.1	97.2	97.2
j	92.0	77.0	59.0	87.8	80.2	48.5	89.2	87.8	89.2
k	82.5	57.0	56.5	63.7	64.9	48.3	65.1	63.7	65.2
l	94.0	68.5	76.5	86.9	75.9	65.1	89.0	88.2	89.5
m	59.0	69.5	60.5	58.3	67.1	49.9	68.5	57.1	65.8
n	100.0	60.5	62.5	98.6	46.9	42.9	98.6	98.6	98.6
Mean	80.8	68.0	63.6	78.9	69.1	57.5	83.8**	80.3	83.8**

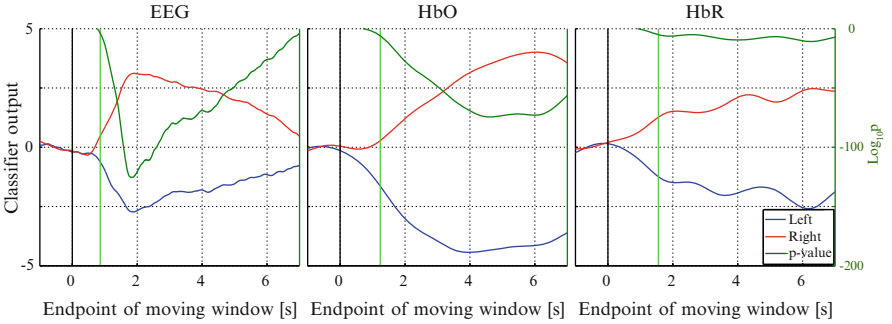

Fig. 3.1 Shows the grand average time-course of classification over all subjects for the individual measurements (EEG – *left*, [HbO] – *center*, [HbR] – *right*). The *red* and *blue lines* depict right and left hand movement imagery, respectively. The *vertical black line* indicates the stimulus onset and the *vertical light green line* indicates the time point at which the means of the classification outputs deviate significantly ($p < 0.001$)

3.2.3 *Multimodal Imaging for Idle-State Detection and Asynchronous BCIs*

To date, a large number of EEG- or NIRS-based asynchronous systems have been proposed. EEG-based asynchronous systems have successfully been applied to real-world applications such as a wheelchair operation [26] as well as rehabilitation devices [15, 57]. Schudlo and Chau developed a real-time NIRS-based online system [52]. They developed and evaluated an online NIRS-BCI which was driven by a mental arithmetic activation task and accommodated an unconstrained rest state. An overall online classification accuracy of 77.4 % demonstrated that mental arithmetic is a potent mental task for driving an online system-paced NIRS-BCI. Other studies have also demonstrated successful applications in online settings of NIRS with a synchronous training paradigm [27, 55].

When comparing the EEG and NIRS modalities in an online asynchronous BCI system, NIRS signals have a critical defect of lower ITR as mentioned in Sect. 3.1. The hemodynamic response delay is the primary limitation preventing NIRS from becoming a fast response BCI system. Because asynchronous BCI systems need to provide a fast response to the users, the ITR problem caused by the nature of the NIRS signal should be solved before developing an EEG+NIRS based asynchronous system. In this study, NIRS signals were used to classify idle and active states (formerly termed 'brain-switch'). We propose to use the inflection points of the NIRS classifiers' output as a transition point of mental states from active to idle and vice versa. This method could reduce the hemodynamic delay significantly, therefore circumvent the time delay of the EEG+NIRS hybrid system.

Data from a previously published study was used [23] (fourteen subjects, right and left hand motor-imagery task) and separated into two parts of training and test data. The first part was used to calculate all types of LDA classifiers, and the second part was employed for performance evaluation. This was done in an offline fashion, however using a pseudo-online technique.

To classify the three mental states (left, right, and idle), a brain switch for detecting the mental state change between active (left or right) and an idle state was created by the NIRS signal which was calculated from the inflection points on the NIRS classifier output.

Three OVR (one-versus rest) classifiers were also calculated from the EEG signal to classify the three mental states individually (right vs. rest, left vs. rest, idle vs. rest). The idea is that the three EEG classifiers are mainly used to classify the three mental states, and the NIRS-based brain switch was used to weigh the output of EEG classifiers. To illustrate further, when the NIRS-based brain switch is on, the outputs of the EEG classifiers containing active classes (right vs. rest and left vs. rest) are raised. Conversely, when the brain switch is off, the output of the EEG classifier containing the idle class (idle vs. rest) is raised. The performance was evaluated in terms of accuracy, sensitivity, specificity, and AUC scores (see Table 3.2).

Our results indicate that the proposed hybrid BCI outperforms the EEG stand-alone system (see also [39]). Paired t-tests between EEG and combined EEG+NIRS

Table 3.2 Shows classification accuracy (Acc), sensitivity (Sen), specificity (Spec) and AUC (all in [%]) for individual subjects as well a their mean, comparing uni-modal EEG well as multimodal EEG+NIRS. * indicates $p < 0.05$, ** indicates $p < 0.01$

	EEG				EEG + NIRS			
Subj	Acc	Sen	Spec	AUC	Acc	Sen	Spec	AUC
a	63.9	26.0	76.8	0.64	60.0	78.6	56.0	0.79
b	76.6	80.0	75.2	0.81	76.7	82.0	74.5	0.81
c	41.9	72.5	33.8	0.61	44.7	73.2	37.1	0.63
d	75.5	62.0	80.1	0.80	75.1	66.4	78.0	0.81
e	55.0	64.3	52.4	0.69	54.7	65.1	51.9	0.69
f	43.7	65.6	37.9	0.62	43.9	66.5	37.9	0.62
g	71.7	78.6	69.3	0.80	73.4	85.6	69.2	0.83
h	63.4	69.7	61.0	0.70	61.8	70.7	58.5	0.69
i	41.9	38.5	43.0	0.50	53.3	54.3	53.1	0.65
j	59.4	54.9	61.6	0.60	64.9	73.9	60.9	0.69
k	32.9	81.5	19.4	0.55	63.9	82.6	58.4	0.78
l	70.0	53.6	77.2	0.68	62.9	71.1	59.7	0.69
m	49.4	54.4	47.7	0.59	61.9	73.5	58.7	0.76
n	79.3	73.6	81.9	0.79	78.3	82.3	76.5	0.82
Mean	58.9	62.5	58.4	0.67	62.5	73.3*	59.3	0.73**

reveal that the accuracy and area under curve (AUC) scores show significantly enhanced classification performance ($p = 0.042$ and $p = 0.013$, respectively). Additionally, sensitivity shows a highly significant result ($p = 0.006$). However, specificity improved but did not show a significant increase. By employing a hybrid-BCI approach [43], the overall performance was significantly enhanced, and some subjects showed remarkably improved accuracy (see subjects VPeaa, VPeak, and VPeam in Table 3.2). While the routines presented here, have been validated in an offline fashion using a pseudo-online technique, our future work will extend our research to a real-time asynchronous hybrid BCI with multimodal visual feedback.

3.3 Discussion and Conclusions

In this chapter, we discussed some of the benefits and limitations of multimodal neuroimaging in the context of BCI. In principle, multimodal data is desirable, since it allows focus on the strengths of the individual methods and furthermore leads to better decoding, which can be attributed to the complementary information content. However, a number of limitations are currently still present. These are reviewed and partial solutions are offered.

When performing multimodal measurements, the setup time is significantly increased. Setup time includes the setup of hardware, but also includes time that is needed recording calibration datasets. Since NIRS does not suffer from volume

conduction, as EEG does, it is a viable target for channel selection, as our results confirm. While in EEG settings, information may get lost, for NIRS only a few channels show similar or even enhanced decoding accuracy compared to whole-head setups.

Subject-independent decoding is a novel technique that allows real-time feedback without the need of recording calibration data. It has been realized for EEG-based BCIs [2, 19–21, 36] and more recently also for NIRS [1]. However, extensions for the combination of multimodal data have not been realized so far. The future will show, whether such systems are feasible.

We have presented some evidence that the real-time integration of NIRS and EEG may be possible to some extent. Detecting the idle-state with EEG is a common and currently insufficiently solved problem in EEG-based BCI systems. To this end we showed that multimodal systems are superior in detecting the idle state, when compared to unimodal EEG. Although not mentioned in this manuscript, EEG-based BCIs also suffer from large perfomance fluctuations. Preliminary results suggest that NIRS-based mental state monitoring is able to detect the slow variations of mental states such as attention or fatigue [8]. Given this state information, state-dependent EEG classfiers can be found to counter this effect [17, 22].

For the future, we envision a real-time asynchronous hybrid BCI system with enhanced decoding rates. This system should retain the responsiveness of EEG-based paradigms and finally require minimal setup time.

Acknowledgements This work was supported by the Brain Korea 21 Plus Program as well as the SGER Grant 2014055911 through the National Research Foundation of Korea funded by the Ministry of Education. This publication only reflects the authors views. Funding agencies are not liable for any use that may be made of the information contained herein. The authors acknowledge the use of some text from the prior publications [23, 38, 63] and thank their co-authors for allowing them to use materials from prior joint publications.

References

1. Abibullaev B, An J, Jin S-H, Lee SH, Moon JI (2013) Minimizing inter-subject variability in fNIRS-based brain–computer interfaces via multiple-kernel support vector learning. Med Eng Phys 35(12):1811–1818
2. Alamgir M, Grosse-Wentrup M, Altun Y (2010) Multitask learning for brain-computer interfaces. In: AISTATS'10: 13th international conference on articial intelligence and statistics. MIT Press, Cambridge, pp 17–24
3. Babiloni F, Cincotti F, Babiloni C, Carducci F, Mattia D, Astolfi L, Basilisco A, Rossini P, Ding L, Ni Y, Cheng J, Christine K, Sweeney J, He B (2005) Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. NeuroImage 24(1):118–131
4. Baillet S, Garnero L, Marin G, Hugonin J-P (1999) Combined MEG and EEG source imaging by minimization of mutual information. IEEE Trans Biomed Eng 46(5):522–534
5. Bauernfeind G, Scherer R, Pfurtscheller G, Neuper C (2011) Single-trial classification of antagonistic oxyhemoglobin responses during mental arithmetic. Med Biol Eng Comput 49(9):979–984

6. Berger H (1929) Über das Elektrenkephalogramm des Menschen. Eur Arch Psychiatry Clin Neurosci 87(1):527–570
7. Bießmann F, Plis S, Meinecke FC, Eichele T, Müller K-R (2011) Analysis of multimodal neuroimaging data. IEEE Rev Biomed Eng 4:26–58
8. Bogler C, Mehnert J, Steinbrink J, Haynes J-D (2014) Decoding vigilance with NIRS. PLoS ONE 9(7):e101729
9. Coyle S, Ward T, Markham C, McDarby G (2004) On the suitability of near-infrared (NIR) systems for next-generation brain–computer interfaces. Physiol Meas 25(4):815
10. Coyle S, Ward T, Markham C (2007) Brain–computer interface using a simplified functional near-infrared spectroscopy system. J Neural Eng 4(3):219
11. Cui X, Bray S, Reiss AL (2010) Speeded near infrared spectroscopy (NIRS) response detection. PLoS ONE 5(11):e15474
12. Dähne S, Bießman F, Meinecke FC, Mehnert J, Fazli S, Müller K-R (2013) Integration of multivariate data streams with bandpower signals. IEEE Trans Multimed 15(5):1001–1013
13. Dähne S, Bießmann F, Samek W, Haufe S, Goltz D, Gundlach C, Villringer A, Fazli S, Müller K-R (2015) Multivariate machine learning methods for fusing functional multimodal neuroimaging data. Proc IEEE 103(9):1507–1530
14. Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci 5(2):162–176
15. Dornhege G, del R. Millán J, Hinterberger T, McFarland D, Müller K-R (eds) (2007) Toward brain-computer interfacing. MIT Press, Cambridge
16. Ehlis A-C, Ringel T, Plichta M, Richter M, Herrmann M, Fallgatter A (2009) Cortical correlates of auditory sensory gating: a simultaneous near-infrared spectroscopy event-related potential study. Neuroscience 159(3):1032–1043
17. Fazli S, Lee S-W (2013) Brain computer interfacing: a multi-modal perspective. J Comput Sci Eng 7(2):132–138
18. Fazli S, Dähne S, Samek W, Bießman F, Müller K-R (2015) Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based Brain-Computer Interfaces. Proc IEEE 103(6):891–906
19. Fazli S, Popescu F, Danóczy M, Blankertz B, Müller K-R, Grozea C (2009) Subject-independent mental state classification in single trials. Neural Netw 22(9):1305–1312
20. S. Fazli, C. Grozea, M. Danoczy, B. Blankertz, F. Popescu, and K.-R. Müller. Subject independent eeg-based bci decoding. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 513–521. Curran Associates, Inc., 2009.
21. Fazli S, Danóczy M, Schelldorfer J, Müller K-R (2011) L1-penalized Linear Mixed-Effects Models for high dimensional data with application to BCI. NeuroImage 56(4):2100–2108
22. S. Fazli, J. Mehnert, J. Steinbrink, and B. Blankertz. Using NIRS as a predictor for EEG-based BCI performance. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4911–4914. IEEE, Aug 2012.
23. Fazli S, Mehnert J, Steinbrink J, Curio G, Villringer A, Müller K-R, Blankertz B (2012) Enhanced performance by a hybrid NIRS–EEG brain computer interface. NeuroImage 59(1):519–529
24. Friston KJ (2009) Modalities, modes, and models in functional neuroimaging. Science 326(5951):399–403
25. Fukunaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic, San Diego
26. Galán F, Nuttin M, Lew E, Ferrez PW, Vanacker G, Philips J, Millán JdR (2008) A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. Clin Neurophysiol 119(9):2159–2169

27. Girouard A, Solovey ET, Jacob RJ (2013) Designing a passive brain computer interface using real time classification of functional near–infrared spectroscopy. Int J Auton Adapt Commun Syst 6(1):26–44
28. Grozea C, Voinescu C, Fazli S (2011) Bristle-sensors – low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. J Neural Eng 8:025008
29. Hoshi Y (2011) Towards the next generation of near-infrared spectroscopy. Philos Trans R Soc A Math Phys Eng Sci 369(1955):4425–4439
30. Huppert T, Hoge R, Diamond S, Franceschini MA, Boas DA (2006) A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. NeuroImage 29(2):368–382
31. Ives J, Warach S, Schmitt F, Edelman R, Schomer D (1993) Monitoring the patient's EEG during echo planar MRI. Electroencephalogr Clin Neurophysiol 87(6):417–420
32. Jobsis FF (1977) Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198(4323):1264–1267
33. Kanoh S, Murayama Y-m, Miyamoto K-i, Yoshinobu T, Kawashima R (2009) A NIRS-based brain-computer interface system during motor imagery: system development and online feedback training. In: Annual international conference of the IEEE engineering in medicine and biology society, Sept 2009 (EMBC 2009). IEEE, Piscataway, pp 594–597
34. Khan B, Wildey C, Francis R, Tian F, Delgado MR, Liu H, MacFarlane D, Alexandrakis G (2012) Improving optical contact for functional near-infrared brain spectroscopy and imaging with brush optodes. Biomed Opt Express 3(5):878–898
35. Khan MJ, Hong MJ, Hong K-S (2014) Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. Front Hum Neurosci 8:244
36. Kindermans P-J, Tangermann M, Müller K-R, Schrauwen B (2014) Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. J Neural Eng 11(3):035005
37. Kocsis L, Herman P, Eke A (2006) The modified Beer-Lambert law revisited. Phys Med Biol 51:N91–N98
38. Lee M-H, Fazli S, Lee S-W (2013) Optimal channel selection based on statistical analysis in high dimensional NIRS data. In: International winter workshop on brain-computer interface (BCI). IEEE, pp 95–97
39. Lee M-H, Fazli S, Mehnert J, Lee S-W (2015) Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI. Pattern Recognit 48(8):2725–2737
40. Lev J (1949) The point biserial coefficient of correlation. Ann Math Stat 20(1):125–126
41. MacKay DJC (2002) Information theory, inference & learning algorithms. Cambridge University Press, New York
42. Matthews F, Pearlmutter BA, Ward TE, Soraghan C, Markham C (2008) Hemodynamics for brain-computer interfaces. IEEE Signal Process Mag 25(1):87–94
43. Pfurtscheller G, Allison BZ, Bauernfeind G, Brunner C, Solis Escalante T, Scherer R, Zander TO, Müller-Putz G, Neuper C, Birbaumer N (2010) The hybrid BCI. Front Neurosci 4:3
44. Pichler BJ, Wehrl HF, Kolb A, Judenhofer MS (2008) Positron emission tomography/magnetic resonance imaging: the next generation of multimodality imaging? Semin Nucl Med 38(3):199–208
45. Piper SK, Krueger A, Koch SP, Mehnert J, Habermehl C, Steinbrink J, Obrig H, Schmitz CH (2014) A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. NeuroImage 85:64–71
46. Popescu F, Fazli S, Badower Y, Blankertz B, Müller K-R (2007) Single trial classification of motor imagination using 6 dry EEG electrodes. PLoS ONE 2(7):e637
47. Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
48. Saad Z, Ropella K, Cox R, DeYoe E (2001) Analysis and use of fMRI response delays. Hum Brain Mapp 13(2):74–93

49. Sannelli C, Dickhaus T, Halder S, Hammer E-M, Müller K-R, Blankertz B (2010) On optimal channel configurations for SMR-based brain–computer interfaces. Brain Topogr 23(2):186–193
50. Schröder M, Lal TN, Hinterberger T, Bogdan M, Hill NJ, Birbaumer N, Rosenstiel W, Schölkopf B (2005) Robust EEG channel selection across subjects for brain-computer interfaces. EURASIP J Appl Signal Process 2005:3103–3112
51. Schroeter ML, Kupka T, Mildner T, Uludağ K, Von Cramon DY (2006) Investigating the post-stimulus undershoot of the BOLD signal–a simultaneous fMRI and fNIRS study. NeuroImage 30(2):349–358
52. Schudlo LC, Chau T (2014) Dynamic topographical pattern classification of multichannel prefrontal NIRS signals: II. Online differentiation of mental arithmetic and rest. J Neural Eng 11(1):016003
53. Ge S, Wang R, Yu D (2014) Classification of four-class motor imagery employing single-channel electroencephalography. PLoS ONE 9(6):e98019
54. Sitaram R, Zhang H, Guan C, Thulasidas M, Hoshi Y, Ishikawa A, Shimizu K, Birbaumer N (2007) Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface. NeuroImage 34(4):1416–1427
55. Solovey ET, Adviser-Jacob RJ (2012) Real-time fNIRS brain input for enhancing interactive systems. Ph.D. thesis, Tufts University
56. Tate RF (1954) Correlation between a discrete and a continuous variable. Point-biserial correlation. Ann Math Stat 25(3):603–607
57. Van De Ville D, Lee S-W (2012) Brain decoding: opportunities and challenges for pattern recognition. Pattern Recognit 45(6):2033–2034
58. Villringer A, Chance B (1997) Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci 20(10):435–442
59. Villringer A, Planck J, Hock C, Schleinkofer L, Dirnagl U (1993) Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci Lett 154(1):101–104
60. Wallois F, Mahmoudzadeh M, Patil A, Grebe R (2012) Usefulness of simultaneous EEG–NIRS recording in language studies. Brain Lang 121(2):110–123
61. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
62. Wriessnegger S, Kurzmann J, Neuper C (2008) Spatio-temporal differences in brain oxygenation between movement execution and imagery: a multichannel near-infrared spectroscopy study. Int J Psychophysiol 67(1):54–63
63. Yeom S-K, Fazli S, Mehnert J, Blankertz B, Steinbrink J, Müller K-R, Lee S-W (2013) Multimodal imaging technique for rapid response brain-computer interface feedback. In: International winter workshop on Brain-Computer Interface (BCI). IEEE, pp 92–94

Chapter 4
Multifrequency Analysis of Brain-Computer Interfaces

Siamac Fazli*, Heung-Il Suk*, Seong-Whan Lee, and Klaus-Robert Müller

Abstract Modern brain computer interfaces (BCI) rely on an extensive use of machine learning and signal processing techniques. This review will focus on an important prerequisite, namely spectral preprocessing. In particular, the optimal usage of multiple frequency features for BCI is discussed in general along with the commonly employed tricks for frequency choice. This is linked to the underlying physiology. Finally, applications of the multifrequency framework are given: (a) to BCI in general and (b) for analysing the BCI illiterates phenomenon.

Keywords BCI • EEG • Spatio-temporal filtering • Filter bank • Bayesian framework • BCI illiteracy

4.1 Introduction

Machine Learning and Signal Processing techniques have been instrumental for the recent advances in Brain Computer Interfaces (BCIs) [11, 14, 31, 56]. Spatio-temporal filtering [9, 22, 25, 33], classification methods [31, 35, 36], projection methods [13, 24, 54] and robust estimation [43, 50] have been actively advanced over the years to develop improved computational methods for better BCI systems. A crucial starting point and input for all of the above methods is a spectral representation of the neural signal as a function of time. Meaningful frequency

*These authors contributed equally.

S. Fazli (✉) • H.-I. Suk • S.-W. Lee
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: fazli@korea.ac.kr; hisuk@korea.ac.kr; swlee@korea.ac.kr

K.-R. Müller
Machine Learning Group, Berlin Institute of Technology, Berlin, Germany
e-mail: klaus-robert.mueller@tu-berlin.de

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_4

bands for noninvasive BCI are the alpha and mu rhythms as well as the beta rhythm [38, 41]. Classical BCIs typically use a single frequency band such as 8–12 hz for covering the mu activity, say for building a motor imagery BCI. However, it has recently been found that a use of multiple frequency bands is highly advantageous yielding better and more robustly performing BCIs [3].

This review aims to explain the importance and advantages of using multiple frequency bands for preprocessing and discusses a number of applications in a BCI context. In a very broad sense using several frequencies captures multimodal aspects of the underlying cognitive processes [12, 20]. Seen from physiology, the use of multiple frequencies allows to fuse different rhythmic components that reflect distinct neural computation that may vary strongly between individuals. As the use of multiple individualized frequencies has become a standard procedure, we will review three common approaches (a) choice of individual frequency bands, (b) the use of intersubject filterbanks and (c) the optimal Bayesian choice of filterbanks for BCI. In addition to an application of multiple frequency techniques to BCI feature construction, we will also discuss the use of multiple frequency techniques for detecting BCI illiteracy.

Both applications are intended to serve as representative examples for these very useful preprocessing procedures.

4.2 Feature Extraction in BCI

The voluntary modulation of sensorimotor rhythms (SMR) presents the basis for SMR-based Brain Computer Interfaces [14, 56]. The μ-rhythm (8–12 Hz) and synchronized components in the β-band (16–22 Hz) are macroscopic idle rhythms that prevail over the postcentral somatosensory cortex and precentral motor cortex, when a given subject is at rest. Imaginations of movements as well as actual movements are known to suppress these idle rhythms contralaterally. This change in neural oscillation is also known as the ERD/ERS[1] effect [41]. With the help of state-of-the-art machine learning and signal processing techniques [7, 14] the ERD/ERS effect can be detected on a single-trial basis and thus be employed for real-time EEG-based BCI feedback.

EEG has high temporal resolution, but poor spatial resolution due to volume conduction. By combining EEG features from multiple spatial locations it is possible to reverse the volume conduction effects to some degree. Such spatial filters for band-power features, computed in frequency bands, can be successfully applied for classification tasks in SMR-based BCI systems. The choice of these frequency bands or *temporal filters* will be discussed in the following.

A popular technique for computing *spatial filters* is termed Common Spatial Pattern (CSP) [9, 22, 25, 42, 50]. This technique allows to focus on spatial locations

[1] Event-related desynchronization/Event-related synchronization.

with the highest ERD/ERS effect. Mathematically a CSP filter is a projection that maximizes the variance of one class, while minimizing the variance of the other class simultaneously. In order to compute the CSP filters W, the algorithm jointly diagonalizes the covariance matrix Σ_i of the trial-concatenated matrix of class $i \in 1, 2$

$$W^\top \Sigma_1 W = D \quad \text{and} \quad W^\top \Sigma_2 W = I - D \tag{4.1}$$

where I is an identity matrix and D is a diagonal matrix with entries d_i ($0 \leq d_i \leq 1$). As eigenvalue d_i is equal to the power ratio of signals of class 1 by class 2 in the corresponding CSP filter (i-th column of matrix W), best discrimination is provided by filters with very high (i.e. near 1) or very low (i.e. near 0) eigenvalues. Typically one would retain projections corresponding to two or three of the highest eigenvalues d_i, i.e., CSP filters for class 1, and projections corresponding to the two or three lowest eigenvalues, i.e., CSP filters for class 2. For a more detailed description of CSP and its application to BCI we would like to refer the reader to [9, 51].

The temporally and spatially filtered data may then be classified by means of simple linear classifiers [11], such as linear discriminant analysis (LDA). LDA assumes the classes to be normally distributed with different means μ_1 and μ_2 but with an identical covariance matrix Σ of full rank. Assuming these quantities to be known, the hyperplane, given by the normal vector $\mathbf{w}_{\text{LDA}}$, can be calculated by:

$$\mathbf{w}_{\text{LDA}} = \Sigma^{-1}(\mu_1 - \mu_2) \tag{4.2}$$

Given that these assumptions hold, the separating hyperplane is *Bayes optimal*.

4.2.1 Estimating a Single Temporal Filter

Traditionally the features for BCIs were set to one specific frequency band (generally the μ-rhythm or broad-band filter) at a single spatial location (in motor related channels, such as 'C3' or 'C4'). However, this approach is often suboptimal. Motor imagery causes the macroscopic idle rhythms to desynchronize, but the exact frequency range and spatial location of this ERD effect is highly subject-dependent and needs to be estimated from calibration data or otherwise. One possibility to do so would be to use a heuristic, based on the point biserial correlation coefficient [9, 30, 32, 48]. The pseudo code of a highly successful heuristic [9] can be obtained in Algorithm 1 (adopted from [9]). The EEG trials X should be spatially filtered by simple spatial filters (such as a Laplacian or a bipolar filter) prior to its use. Best results are achieved when using only a few channels from motor-related areas: e.g., to choose $C = \{c_1, c_2, c_3\}$ with c_i being one from each area of the left hand, right hand and feet with max $\sqrt{\sum_f (\text{score}_c(f))^2}$.

Algorithm 1 Selection of a discriminative frequency band (Adopted from [9])

Require: Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.

1: $\mathrm{dB}_c(f, i) \leftarrow$ log band-power of $X_{(c,i)}$ at frequency f (f from 5 to 35 Hz)
2: $\mathrm{score}_c(f) \leftarrow \mathrm{corrcoef}\,(\mathrm{dB}_c(f, i),\ y_i)_i$
3: $f_{\max} \leftarrow \mathrm{argmax}_f \sum_{c \in C} \mathrm{score}_c(f)$
4: $\mathrm{score}^*_c(f) \leftarrow \begin{cases} \mathrm{score}_c(f) & \text{if } \mathrm{score}_c(f_{\max}) > 0 \\ -\mathrm{score}_c(f) & \text{otherwise} \end{cases}$
5: $\mathrm{fscore}(f) \leftarrow \sum_{c \in C} \mathrm{score}^*_c(f)$
6: $f^*_{\max} \leftarrow \mathrm{argmax}_f\, \mathrm{fscore}(f)$
7: $f_0 \leftarrow f^*_{\max};\ f_1 \leftarrow f^*_{\max}$
8: **while** $\mathrm{fscore}(f_0 - 1) \geq \mathrm{fscore}(f^*_{\max}) * 0.05$ **do**
9: $\quad f_0 \leftarrow f_0 - 1$
10: **while** $\mathrm{fscore}(f_1 + 1) \geq \mathrm{fscore}(f^*_{\max}) * 0.05$ **do**
11: $\quad f_1 \leftarrow f_1 + 1$
12: **return** frequency band $[f_0, f_1]$

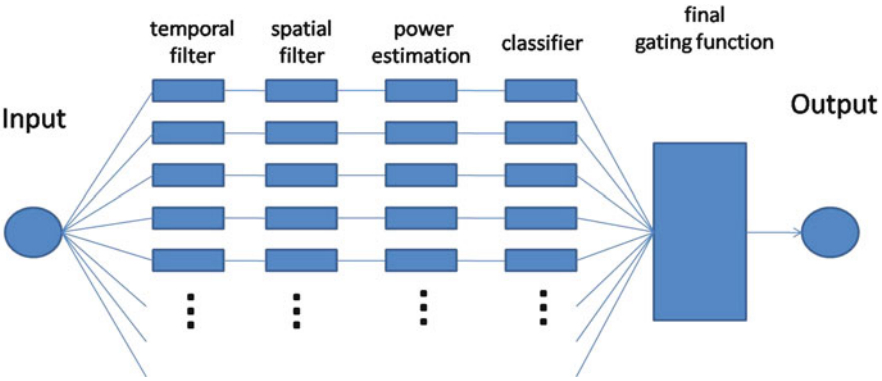

Fig. 4.1 Flowchart of parallel multiple frequency band processing

4.2.2 Combination of Multiple Temporal Filters

While estimating a single subject-dependent frequency band requires minimal computation time and leads to very stable results for most subjects, for some subjects the combination of multiple frequencies is favourable [2, 4, 17]. Figure 4.1 show a flowchart of the necessary parallel processing for multiple frequency bands.

The filter bank CSP (FBCSP) algorithm [2] computes CSP features in various frequency bands and applies a feature selection method in order to identify the most informative bands. Several strategies have been proposed for this task. Among the most common is the mutual information-based best individual feature algorithm (MIBIF) [4]. An extensive evaluation of this algorithm on two BCI data competition

IV data sets has recently demonstrated [4] that a hybrid BCI approach based on the integration of multiple frequency bands significantly outperforms single-feature systems in terms of classification accuracy. For this approach a whole range of predefined filters are employed and only their optimal combination needs to be found.

The combination of predefined multiple frequency bands ultimately leads to subject-independent BCI decoding, so-called *zero-training BCI* [1, 18, 19, 21, 26], since now predefined temporal filters can be derived from prior neurophysiological knowledge and the need of a calibration session to estimate such filters is alleviated by finding their optimal combination (see *final gating function* in Fig. 4.1) in an out-of-sample way, i.e. by obtaining this information from *other subjects' data*.

4.2.3 Bayesian Combination of Multiple Temporal Filters

The previous work presented that the use of features from multiple frequency bands helps improve the classification accuracy [2, 49, 57], where the features are highly dependent on the frequency bands under consideration. To our best knowledge, the existing methods mostly predefined the equally-spaced frequency bands in the range between 5 and 30 Hz, including both μ- and β-bands. It is noteworthy that from a machine learning perspective, it is important to find the optimal frequency bands, from which we can extract features discriminative between target classes, thus boosting BCI performance. To tackle this problem, we formulated a Bayesian framework in which the optimal frequency bands are determined in a probabilistic manner. Specifically, we regarded a frequency band as a random variable and tried to estimate the unknown probability density function (*pdf*) based on which we can find class-discriminative frequency bands and features, thus helping to perform multifrequency analysis.

Here, we briefly describe our method proposed in [45]. Let $\mathbf{B} \equiv (b^1, b^2)$ denote a continuous random variable for a frequency band, where b^1 and b^2 are, respectively, the start and the end point of a frequency band ($b^1 < b^2$). We define the probability of a specific frequency band $\mathbf{b}$, $p(\mathbf{b})$, as the chance that the $\mathbf{b}$ bandpass-filtered signals can be used to correctly classify between classes.

Since we are presumably uncertain about the discriminative frequency band, we encode this uncertainty as a prior distribution $p(\mathbf{B})$ over a random variable $\mathbf{B}$. Given a set of single-trial EEGs $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{D}$ and the corresponding class labels $\mathbf{\Omega} = \{\omega_i\}_{i=1}^{D}$, where D is the number of trials, we can compute the posterior *pdf*, $p(\mathbf{B}|\mathbf{X}, \mathbf{\Omega})$, by the Bayes rule as follows

$$p(\mathbf{B}|\mathbf{X}, \mathbf{\Omega}) = \frac{p(\mathbf{X}, \mathbf{\Omega}|\mathbf{B})p(\mathbf{B})}{p(\mathbf{X}, \mathbf{\Omega})}. \tag{4.3}$$

The prior, $p(\mathbf{B})$, describes the relative probabilities of different states, i.e., frequency bands, in which single-trial EEGs pertinent to motor imageries are correctly

discriminated. The term $p(\mathbf{X}, \mathbf{\Omega}|\mathbf{B})$ is a likelihood function. If the hypothesis $\mathbf{B}$, i.e., the frequency band, were true, this term indicates the probability that the single-trial EEGs $\mathbf{X}$ in conjunction with the class labels $\mathbf{\Omega}$ would have been available to support it. The posterior distribution $p(\mathbf{B}|\mathbf{X}, \mathbf{\Omega})$ defines the probability of frequency band $\mathbf{B}$ being true, given the observations of $\mathbf{X}$ and $\mathbf{\Omega}$. Thus, it indicates the relative likelihood of the single-trial EEGs $\mathbf{X}$ being correctly classified into $\mathbf{\Omega}$ by $\mathbf{B}$ bandpass filtering along with the ensuing processes of spatial filtering and feature extraction as described in Sect. 4.2.

Given a frequency band $\mathbf{B}$ and raw EEG signals $\mathbf{X}$, the bandpass-filtered signals $\mathbf{Z}$ can be obtained deterministically. Hence, the likelihood $p(\mathbf{X}, \mathbf{\Omega}|\mathbf{B})$ and the evidence $p(\mathbf{X}, \mathbf{\Omega})$ are equal to $p(\mathbf{Z}, \mathbf{\Omega}|\mathbf{B})$ and $p(\mathbf{Z}, \mathbf{\Omega})$, respectively. We can then rewrite Eq. (4.3) by substituting the bandpass-filtered signals $\mathbf{Z}$ in places of the raw EEGs $\mathbf{X}$ as follows

$$p(\mathbf{B}|\mathbf{Z}, \mathbf{\Omega}) = \frac{p(\mathbf{Z}, \mathbf{\Omega}|\mathbf{B})p(\mathbf{B})}{p(\mathbf{Z}, \mathbf{\Omega})}. \tag{4.4}$$

The posterior $p(\mathbf{B}|\mathbf{Z}, \mathbf{\Omega})$ represents all the knowledge about $\mathbf{B}$ that is deducible from the bandpass-filtered single-trials $\mathbf{Z}$ and the corresponding class labels $\mathbf{\Omega}$.

Note that a spatial filter $\mathbf{W}$ can be found from $\mathbf{Z}$ via an CSP algorithm,[2] in which $\mathbf{W}$ is analytically obtained by computing a generalized eigenvector problem. As described above, a feature vector is extracted by computing simple matrix multiplication between $\mathbf{Z}$ and $\mathbf{W}$ and the second-order statistics followed by a monotonically increasing logarithmic function. It means that the posterior $p(\mathbf{B}|\mathbf{Z}, \mathbf{\Omega})$ can be indirectly estimated from $p(\mathbf{B}|\mathbf{F}, \mathbf{\Omega})$, where $\mathbf{F} = \log\left[\mathrm{var}\left(\mathbf{W}^{\dagger}\mathbf{Z}\right)\right]$ and † denotes a matrix transpose, without losing information in the data. Therefore, we can rewrite Eq. (4.4) as follows with the feature vector set $\mathbf{F}$ extracted from the spatially filtered signals of $\mathbf{Z}$

$$\begin{aligned} p(\mathbf{B}|\mathbf{Z}, \mathbf{\Omega}) &\triangleq p(\mathbf{B}|\mathbf{F}, \mathbf{\Omega}) \\ &= \frac{p(\mathbf{F}, \mathbf{\Omega}|\mathbf{B})p(\mathbf{B})}{p(\mathbf{F}, \mathbf{\Omega})} \end{aligned} \tag{4.5}$$

where $p(\mathbf{F}, \mathbf{\Omega}) = \int_{\mathbf{B}} p(\mathbf{F}, \mathbf{\Omega}|\mathbf{B})p(\mathbf{B})d\mathbf{B}$. Thus, our goal of finding the optimal spatio-spectral filter for discriminative feature extraction, ultimately improving classification accuracy, can be defined as estimation of the posterior *pdf* $p(\mathbf{B}|\mathbf{F}, \mathbf{\Omega})$ in Eq. (4.5), for which a sequential Monte Carlo method [15] was used in [45].

[2] Basically, any variants of the standard CSP [8, 16, 43, 44] can be used for this.

4.3 Applications

The physiological patterns induced by ERD/ERS effects exhibit high variability across subjects and even trials for the same subjects [42]. Furthermore, the EEG electrodes measure the superimposed signals that originate from various sources in the brain and the EEG signals are generally contaminated with artifacts and noises that can cause performance degradation in classification. Therefore, there has been considerable interest in finding the ERD/ERS-related frequency band(s) and the class-discriminative spatial patterns within the BCI community. The estimation of appropriate subject-dependent spatial and temporal filters are of paramount importance for high-speed real-time SMR-based feedback sessions. In this regard, the user-specific calibration of a BCI has been one of the main steps for practical application. Among the many realized applications are some, which are intended for users with disabilities, such as spelling devices [55], wheelchair operation [23] or grabbing object by means of an external robot arm [40], but also some gaming applications, such as pinball [47], brain pong [28, 37] and other games [27, 39] have been realized.

4.3.1 Illiteracy Prediction

The successful control of a SMR-based BCI system varies greatly among subjects [10] and for approximately 20 % of subjects the decoding accuracy does not reach a sufficient threshold and it is therefore not possible for users to control an application in a meaningful way. A technique which partly solves the *BCI-illiteracy phenomenon* has already been proposed [52, 53], but this topic requires further attention from the scientific community. Recently a method for predicting BCI illiteracy has been suggested [10]. In this method the spectral components of Laplace-filtered, motor-related resting-state EEG are examined and features extracted for the successful prediction of SMR-based real-time feedback performance. The results show a correlation of $r = 0.53$ between the predictor and BCI performance on a large-scale database of 80 BCI-naive participants. These results contribute to counteract the BCI illiteracy effect in the sense that potentially non-successful subjects can be identified at the very beginning of a session and to these subjects operant conditioning procedures [6, 29, 34] could be applied in a pre-training session to specifically enhance the power of the SMR idling rhythm.

4.3.1.1 Bayesian Multifrequency Analysis for BCI Illiterate Detection

Thanks to the data-driven multifrequency analysis of our framework described in Sect. 4.2.3, we can extract a subject-specific filter distribution that can be analyzed to gain a better understanding of individual differences of BCI users. To identify BCI

illiterates, we studied the prediction of a subject's future BCI performance based on resting-state EEG data acquired prior to a BCI session. Using our framework, we estimated spatial patterns and clustered subjects according to the patterns corresponding to the extracted filter characteristics. We then analyzed the resulting grouping in order to gain a better physiological understanding why some subjects perform better than others and what the characteristics of subjects with BCI-inability could be [46].

Using only 3-Laplacian channels, our predictions yielded a correlation coefficient of 0.6 with the performance later seen in the actual BCI feedback session, which is favourably comparative with previous results [10]. A clustering of the resulting spatial patterns showed interesting task-independent physiological characteristics discriminative for good and bad BCI performers. In line with [10], a strong idle μ-rhythm is indicative for a high BCI accuracy. Nevertheless, more complex indicators can be observed in 5 groups each with different characteristic peak frequency in the μ-band and possible further modulation of the β-band. The spectral characteristics in a subject's resting period thus appears already indicative whether the subjects will belong to a good, mediocre, or under-performing group in BCI usage.

4.4 When Are Combinations Useful?

The combination of multiple frequency bands has previously been shown to be particularly useful, for subjects where the statistical differences of frequencies are not prominent. These statistical differences can be estimated by the point-biserial correlation coefficient. In these cases heuristics, which estimate statistical scores, can be error prone. In addition when no calibration data is available, predefined filter banks work better as compared to single broad band filters or filters focussing on the μ-rhythm [2, 17]. However, some care should be taken, when considering multiple temporal filters. Adding features, which do not contain task-related information or redundant features can harm decoding accuracy. Therefore information theoretic measures need to be calculated, to find out whether the extra information is useful [2, 20].

4.5 Conclusion

Extracting meaningful features that reflect the spectral properties of EEG well is a prerequisite for all following analysis steps in BCI or general cognitive neuroscience experiments. It is well known in physiology that one frequency band alone is oftentimes insufficient to capture a complex cognitive process.

We have therefore reviewed techniques that allow the usage of multiple frequencies, starting from single frequency analysis, to usage of two frequencies (e.g. alpha

and beta bands), we finally discuss the recent optimal Bayesian modelling of the full spectrum. After having established the machinery, we discuss its application for generic BCIs and also its use for the prediction of illiterates. While the use of multifrequency analysis has increased the accuracy and usability of SMR-based BCIs, there are still a number of open challenges, which need to be addressed.

More generally speaking multi-frequency techniques are one instance of data fusion methods [5, 12, 20], where multiple sources or modes that carry information are harvested. Note that research aiming to optimally fuse multimodal information has recently received wide attention not only in neuroimaging but also in the sciences and also in the engineering disciplines.

Acknowledgements This work was supported by the Brain Korea 21 Plus Program as well as the SGER Grant 2014055911 through the National Research Foundation of Korea funded by the Ministry of Education. This publication only reflects the authors views. Funding agencies are not liable for any use that may be made of the information contained herein. The authors acknowledge the use of some text from the prior publications [45, 46] and thank their co-authors for allowing them to use materials from prior joint publications.

References

1. Alamgir M, Grosse-Wentrup M, Altun Y (2010) Multitask learning for brain-computer interfaces. In: AISTATS'10: 13th international conference on articial intelligence and statistics, May 2010. MIT Press, Cambridge, pp 17–24. http://jmlr.csail.mit.edu/proceedings/papers/v9/
2. Ang KK, Chin ZY, Zhang H, Guan C (2008) Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: IEEE international joint conference on neural networks (IJCNN). IEEE, Piscataway, pp 2390–2397
3. Ang KK, Chin ZY, Wang C, Guan C, Zhang H (2012) Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front Neurosci 6:00039. http://www.frontiersin.org/Journal/Abstract.aspx?s=763&name=neuroprosthetics&ART_DOI=10.3389/fnins.2012.00039
4. Ang KK, Chin ZY, Zhang H, Guan C (2012) Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. Pattern Recognit 45(6):2137–2144
5. Bießmann F, Plis SM, Meinecke FC, Eichele T, Müller KR (2011) Analysis of multimodal neuroimaging data. IEEE Rev Biomed Eng 4:26–58
6. Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H (1999) A spelling device for the paralysed. Nature 398:297–298
7. Blankertz B, Dornhege G, Krauledat M, Müller KR, Curio G (2007) The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. Neuroimage 37(2):539–550. http://dx.doi.org/10.1016/j.neuroimage.2007.01.051
8. Blankertz B, Kawanabe M, Tomioka R, Hohlefeld FU, Nikulin V, Müller KR (2008) Invariant common spatial patterns: alleviating nonstationarities in brain-computer interfacing. In: Advances in neural information processing systems. MIT Press, Cambridge, p 2008
9. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 25(1):41–56. http://dx.doi.org/10.1109/MSP.2008.4408441
10. Blankertz B, Sannelli C, Halder S, Hammer EM, Kübler A, Müller KR, Curio G, Dickhaus T (2010) Neurophysiological predictor of SMR-based BCI performance. Neuroimage 51(4):1303–1309. http://dx.doi.org/10.1016/j.neuroimage.2010.03.022

11. Blankertz B, Lemm S, Treder MS, Haufe S, Müller KR (2011) Single-trial analysis and classification of ERP components – a tutorial. Neuroimage 56:814–825. http://dx.doi.org/10.1016/j.neuroimage.2010.06.048
12. Fazli S, Dähne S, Samek W, Bießmann F, Müller K-R (2015) Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based Brain-Computer Interfaces. Proc IEEE 103(6):891–906
13. Dähne S, Bießmann F, Meinecke FC, Mehnert J, Fazli S, Müller KR (2013) Integration of multivariate data streams with bandpower signals. IEEE Trans Multimed 15(5):1001–1013. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6472075&tag=1
14. Dornhege G, del R. Millán J, Hinterberger T, McFarland D, Müller KR (eds)(2007) Toward brain-computer interfacing. MIT Press, Cambridge
15. Doucet A, De Freitas N, Gordon N (eds) (2001) Sequential Monte Carlo methods in practice. Springer, New York
16. Falzon O, Camilleri K, Muscat J (2012) The analytic common spatial patterns method for EEG-based BCI data. J Neural Eng 9(4):045009
17. Fazli S, Grozea C, Danóczy M, Blankertz B, Müller KR, Popescu F (2008) Ensembles of temporal filters enhance classification performance for ERD-based BCI systems. In: 4th international brain-computer interface workshop and training course. Verlag der Technischen Universität Graz, Graz
18. Fazli S, Popescu F, Danóczy M, Blankertz B, Müller KR, Grozea C (2009) Subject-independent mental state classification in single trials. Neural Netw 22(9):1305–1312. http://dx.doi.org/10.1016/j.neunet.2009.06.003
19. Fazli S, Grozea C, Danóczy M, Blankertz B, Popescu F, Müller K (2009) Subject independent EEG-based BCI decoding. In: Bengio Y, Schuurmans D, Lafferty J, Williams CKI, Culotta A (eds) Advances in neural information processing systems 22. MIT Press, Cambridge, pp 513–521
20. Dähne S, Bießmann F, Samek W, Haufe S, Goltz D, Gundlach C, Villringer A, Fazli S, Müller K-R (2015) Multivariate machine learning methods for fusing functional multimodal neuroimaging data. Proc IEEE 103(9):1507–1530
21. Fazli S, Danóczy M, Schelldorfer J, Müller KR (2011) L1-penalized Linear Mixed-Effects Models for high dimensional data with application to BCI. Neuroimage 56(4):2100–2108. http://www.sciencedirect.com/science/article/pii/S1053811911003405
22. Fukunaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic, Boston
23. Galán F, Nuttin M, Lew E, Ferrez P, Vanacker G, Philips J, Millán J (2008) A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. Clin Neurophysiol 119(9):2159–2169
24. Hyvärinen A, Karhunen J, Oja E (2004) Independent component analysis, vol 46. Wiley, New York
25. Koles ZJ (1991) The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroencephalogr Clin Neurophysiol 79(6):440–447
26. Krauledat M, Tangermann M, Blankertz B, Müller KR (2008) Towards zero training for brain-computer interfacing. PLoS ONE 3(8):e2967. http://dx.doi.org/10.1371/journal.pone.0002967
27. Krepki R, Blankertz B, Curio G, Müller KR (2007) The Berlin brain-computer interface (BBCI)–towards a new communication channel for online control in gaming applications. Multimed Tools Appl 33(1):73–90
28. Krepki R, Blankertz B, Curio G, Müller KR (2007) The Berlin brain-computer interface (BBCI): towards a new communication channel for online control in gaming applications. J Multimed Tools Appl 33(1):73–90. http://dx.doi.org/10.1007/s11042-006-0094-3
29. Kübler A, Nijboer F, Mellinger J, Vaughan TM, Pawelzik H, Schalk G, McFarland DJ, Birbaumer N, Wolpaw JR (2005) Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. Neurology 64(10):1775–1777
30. Lemm S, Blankertz B, Curio G, Müller KR (2005) Spatio-spectral filters for improving the classification of single trial EEG. IEEE Trans Biomed Eng 52:1541–1548

31. Lemm S, Blankertz B, Dickhaus T, Müller KR (2011) Introduction to machine learning for brain imaging. Neuroimage 56:387–399. http://dx.doi.org/10.1016/j.neuroimage.2010.11.004
32. Lev J (1949) The point biserial coefficient of correlation. Ann Math Stat 20(1):125–126
33. Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B (2007) A review of classification algorithms for EEG-based brain-computer interfaces. J Neural Eng 4:R1–R13
34. McFarland DJ, McCane LM, Wolpaw JR (1998) EEG-based communication and control: short-term role of feedback. IEEE Trans Rehabil Eng 6(1):7–11
35. Müller KR, Mika S, Rätsch G, Tsuda K, Schölkopf B (2001) An introduction to kernel-based learning algorithms. IEEE Neural Netw 12(2):181–201
36. Müller KR, Anderson CW, Birch GE (2003) Linear and non-linear methods for brain-computer interfaces. IEEE Trans Neural Syst Rehabil Eng 11(2):165–169
37. Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 167(1):82–90. http://dx.doi.org/10.1016/j.jneumeth.2007.09.022
38. Niedermeyer E, da Silva FL (2005) Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, Philadelphia
39. Nijholt A (2009) BCI for games: a 'state of the art' survey. In: Entertainment computing-ICEC 2008. Springer, Berlin, pp 225–228
40. Onose G, Grozea C, Anghelescu A, Daia C, Sinescu C, Ciurea A, Spircu T, Mirea A, Andone I, Spânu A, Popescu C, Mihaescu S, Fazli S, Danoczy M, Popescu F (2012) On the feasibility of using motor imagery EEG-based brain–computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. Spinal Cord 50(8):599–608. http://www.nature.com/sc/journal/vaop/ncurrent/full/sc201214a.html, open access
41. Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110(11):1842–1857
42. Ramoser H, Müller-Gerking J, Pfurtscheller G (1998) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
43. Samek W, Kawanabe M, Müller KR (2014) Divergence-based framework for common spatial patterns algorithms. IEEE Rev Biomed Eng 7:50–72
44. Shi LC, Li Y, Sun RH, Lu BL (2011) A sparse common spatial pattern algorithm for brain-computer interface. In: Lu BL, Zhang L, Kwok J (eds) Neural information processing. Lecture notes in computer science, vol 7062. Springer, Berlin/Heidelberg, pp 725–733
45. Suk HI, Lee SW (2013) A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces. IEEE Trans Pattern Anal Mach Intell 35(2):286–299
46. Suk HI, Fazli S, Mehnert J, Müller KR, Lee SW (2014) Predicting BCI subject performance using probabilistic spatio-temporal filters. PLoS One 9(2):e87056
47. Tangermann M, Krauledat M, Grzeska K, Sagebaum M, Blankertz B, Vidaurre C, Müller KR (2009) Playing pinball with non-invasive BCI. In: Koller D, Schuurmans D, Bengio Y, Bottou L (eds) Advances in neural information processing systems 21, 8–11 Dec 2008. MIT Press, Cambridge, pp 1641–1648
48. Tate RF (1954) Correlation between a discrete and a continuous variable. Point-biserial correlation. Ann Math Stat 25(3):603–607
49. Thomas K, Guan C, Lau C, Vinod A, Ang K (2009) A new discriminative common spatial pattern method for motor imagery brain-computer interfaces. IEEE Trans Biomed Eng 56(11):2730–2733
50. Tomioka R, Müller KR (2010) A regularized discriminative framework for EEG analysis with application to brain-computer interface. Neuroimage 49:415–432 http://dx.doi.org/10.1016/j.neuroimage.2009.07.045
51. Tomioka R, Dornhege G, Nolte G, Aihara K, Müller K-R (2006) Optimizing spectral filters for single trial EEG classification. In: Proceedings of DAGM. Lecture notes in computer science, vol 4174. Springer, Berlin/Heidelberg, pp 414–423

52. Vidaurre C, Sannelli C, Müller KR, Blankertz B (2011) Co-adaptive calibration to improve BCI efficiency. J Neural Eng 8(2):025009 (8pp). http://dx.doi.org/10.1088/1741-2560/8/2/025009
53. Vidaurre C, Sannelli C, Müller KR, Blankertz B (2011) Machine-learning based co-adaptive calibration. Neural Comput 23(3):791–816. http://dx.doi.org/10.1162/NECO_a_00089
54. von Bünau P, Meinecke FC, Király F, Müller KR (2009) Finding stationary subspaces in multivariate time series. Phys Rev Lett 103:214101
55. Williamson J, Murray-Smith R, Blankertz B, Krauledat M, Müller KR (2009) Designing for uncertain, asymmetric control. Interaction design for brain-computer interfaces. Int J Hum Comput Stud 67(10):827–841. http://dx.doi.org/10.1016/j.ijhcs.2009.05.009
56. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
57. Zhang H, Chin Z, Ang K, Guan C, Wang C (2011) Optimum spatio-spectral filtering network for brain-computer interface. IEEE Trans Neural Netw 22(1):52–63

Part II
Cognitive- and Neural-Rehabilitation Engineering

Chapter 5
Current Trends in Memory Implantation and Rehabilitation

Hyun Jae Jang, Sahn Woo Park, and Jeehyun Kwag

Abstract Hippocampus is believed to be the brain region critical for memory storage and recall. Damage to the hippocampus by lesions or neurodegenerative diseases such as Alzheimer's disease could lead to memory deficits. However, there is yet no treatment method available. Direct deep-brain stimulation (DBS) of the hippocampus has been attempted in an effort to find a treatment method for memory dysfunction and Alzheimer's disease in the last few decades but with limited success. Recently, a novel approach has been developed where an implantation of a very large scale integration (VLSI) microchip containing a biomimetic computational model could act as an artificial bridge to replace the damaged hippocampal circuit *in vivo*. Here, we discuss the memory implantation techniques; from the conventional DBS method to the current memory implantation technology using an artificial neural microchip. Furthermore, we propose future directions towards the development of a physiologically realistic memory implantation chip design that could enhance the performance of the memory implant and be used for the treatment of memory-related neurodegenerative diseases.

Keywords Memory implantation • Memory rehabilitation • Hippocampus • Synaptic plasticity • Neuromorphic chip

5.1 Introduction

Alzheimer's disease is a neurodegenerative disorder where Aβ plaque-associated degeneration of hippocampal neurons leads to cognitive deficits and irreversible memory impairments [1, 2]. Although about 11 % of population over the age of 65 in the United States suffer from Alzheimer's disease [3], current treatments such as the use of medication or cognitive rehabilitation are mainly focused on slowing down the progression of the symptoms and they cannot cure memory deficits caused

H.J. Jang • S.W. Park • J. Kwag (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: jkwag@korea.ac.kr

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_5

by hippocampal damages [4]. Over the last few decades, many studies have tried to find a way to enhance the damaged hippocampal function by artificially stimulating the damaged hippocampal region [5–7] or implanting a microchip that can substitute the missing circuits of the damaged hippocampus [8]. Recently, implantation of a microchip containing a biomimetic computational model was suggested as a novel treatment for the hippocampal damage [9, 10]. In this chapter, we discuss the current trends in memory enhancement and implantation using artificial neural signal stimulation and biomimetic modeling of hippocampal circuit. After outlining the limitations of the aforementioned technologies, we propose future directions in designing an implantable memory chip that incorporates the physiologically realistic computational model of hippocampal circuit, which could have a potential for the treatment of memory-related neurodegenerative diseases.

5.1.1 Memory and Hippocampus

Hippocampus is believed to be the region that is critical for the formation of memory [11]. Hippocampus is located in the medial temporal lobe underneath the cortical surface and the hippocampal circuit consists of hippocampal sub-fields such as Cornus Ammonis (CA) 1, CA3, dentate gyrus (DG), and subiculum (Fig. 5.1a). Hippocampus receives multi-modal sensory inputs from the entorhinal cortex through the perforant path to granule cells in DG and dentate granule cells synapse onto CA3 pyramidal neurons, which in turn synapse onto CA1 pyramidal neurons, forming a tri-synaptic circuit [12]. The final output of the hippocampus arises from CA1 pyramidal neurons that send their output to the entorhinal cortex, completing a recurrent connection between the entorhinal cortex and the hippocampus (Fig. 5.1a). A landmark clinical study on patient H.M. provided proofs for the statement that this simple hippocampal circuitry is crucial in memory formation [11, 13]. Patient H.M., who underwent bilateral hippocampal lobotomy to treat severe epilepsy, had serious anterograde amnesia, not being able to form new memory while having intact past memories and other cognitive abilities [14, 15]. However, how does the hippocampal circuit support the formation of memory? Synaptic plasticity, the change of synaptic efficacy between neurons, is believed to be the most probable substrate for memory [16]. It has been shown that the increase in synaptic efficacy, called long-term potentiation (LTP), or the decrease in synaptic efficacy, called long-term depression (LTD), can be induced by high-frequency stimulation or low-frequency stimulation of presynaptic neurons, respectively [16–22]. In addition, precisely timed and ordered presynaptic spikes paired with postsynaptic spikes occurring within a time window of ± 10 ms have been shown to induce synaptic plasticity, called spike timing-dependent plasticity (STDP) [18, 19]. That is, different neural activity patterns among sparsely activated hippocampal neural population seem to hold a key in inducing different polarity of the synaptic plasticity that is important for hippocampal memory storage [23].

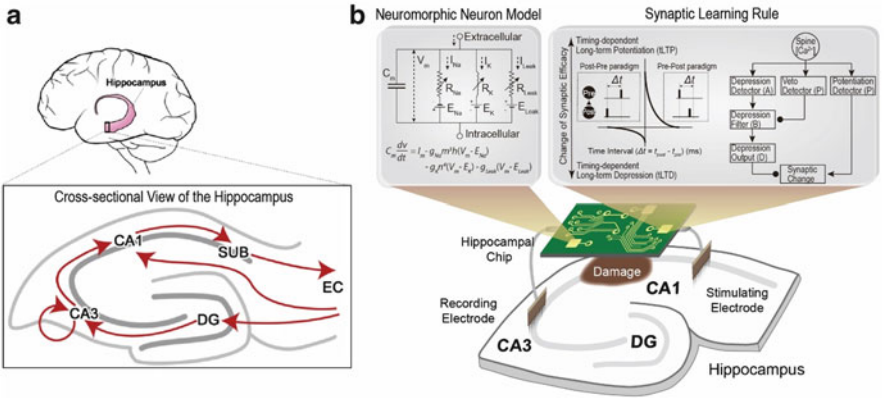

Fig. 5.1 Schematic diagram of the hippocampus and neuromorphic memory implantation technique. (**a**) The location of the hippocampus in the brain (*top*) and a cross-sectional view of the hippocampus (*bottom*). *Red arrows* show the path of the neural signal flow in the hippocampus. *Inset*: entorhinal cortex (*EC*), dentate gyrus (*DG*), Cornus Ammonis 3 (*CA3*), Cornus Ammonis 1 (*CA1*), subiculum (*SUB*). (**b**) Conceptual schematic design of the neuromorphic hippocampal network model for memory implantation. The artificial hippocampal chip containing the neuromorphic hippocampal neuron model and synaptic learning rule (*top*) [59]. Pre-synaptic signals recorded from CA3 neurons of the hippocampus is transmitted to the post-synaptic CA1 neurons through the artificial hippocampal chip (*bottom*), bypassing the damaged hippocampal region

5.2 Memory Implantation and Rehabilitation

Then, how can we enhance or implant memory in the damaged hippocampus? Based on the mechanisms of hippocampal synaptic plasticity, broadly, two different approaches can be made: the one is to artificially generate the neural activity conducive to induce hippocampal synaptic plasticity and the other is to replace the damaged neural circuit directly.

5.2.1 Memory Implantation and Rehabilitation Using Deep-Brain Stimulation (DBS)

Following the finding that the hippocampus in the temporal lobe is critical in memory function [11] and with the development of electrical brain stimulation method [24, 25], many studies have focused on enhancing memory functions by artificially generating neural activities in the hippocampus using electrical stimulation [17, 26, 27]. The very first attempt was made by Wilder Penfield. When he electrically stimulated the temporal lobe in epileptic patients, some of the patients reported to have experienced reactivation of past memories [28]. From then on, many studies have attempted direct activation of the temporal lobe using

direct deep-brain stimulation (DBS) technique to enhance or implant memory [29, 30]. In DBS, an electrode plate is implanted into a specific brain region to generate electrical stimulation. Long-period, chronic electrical stimulation by DBS directly alters the neural activity artificially. Especially, high-frequency DBS has been proven to be effective in treating depression [31, 32] and movement-related neurodegenerative disorders such as Parkinson's disease [24, 25]. Recently, DBS applied to the entorhinal cortex has been reported to enhance spatial memory in human patients with epilepsy [33]. However, DBS directly applied to the hippocampus had little effect on memory enhancement. In many cases, it even caused memory disruption in human brains [5–7], possibly due to the inappropriately large stimulation strength causing continuous over-activation of large population of hippocampal neurons, consequently disrupting their normal computational capacity. That is, hippocampal DBS as a memory prosthetics has a major limitation in that there is a poor understanding of the underlying mechanism of DBS-induced neural activity change due to the complex and intricate brain activity [34]. Thus, careful design of stimulation strength and temporal pattern that reflects the physiologically realistic neural activity patterns of hippocampal neurons is critical in modulating memory using electrical stimulation [35]. Recently, a novel memory implantation technology has been proposed whereby an implantation of an electrical chip that can artificially predict and stimulate the neural activities of the hippocampus based on the *in vivo*-recorded neural activities, acting as an artificial bridge to replace the damaged hippocampal region [9]. Such stimulation of predicted spike trains onto the hippocampal circuit could enhance memory function in rodents [9, 10] and non-human primates with hippocampal damage [36], demonstrating that memory prosthesis is indeed possible.

5.2.2 Memory Implantation and Rehabilitation Using Multi-Input and Multi-Output Model

The pioneering study on memory implantation using the implantation of biomimetic chip was first performed by Berger and his colleagues [9, 10]. They aimed to develop a memory prosthetic device that is able to replace the impaired hippocampal memory circuitry. Firstly, they created an impaired hippocampal neural circuit, by making a chemical lesion in the synaptic pathway between hippocampal CA3 and CA1 regions of rodents using a glutamatergic synapse blocker injected *in vivo* [9]. Rodents with the impaired CA3-CA1 synaptic connections showed poor performance on the working memory task called a delayed non-matched sample (DNMS) task [37, 38]. During the DNMS task, a rodent is trained to remember the chosen option with a delay between selection and retrieval. The rodent is presented with a single lever randomly in either the left or the right position, and the rodent selects to press one of the levers, called the "sample" lever. After random time delays (1–30 s) following the "sample" lever selection, now two levers were presented in both the left and the right position again and the rodent would get a reward for

pressing the opposite lever to the "sample" lever. In the rodent with chemical lesion in the CA3-CA1 synapse, the correction rate of the DNMS task was decreased by 50 % compared to that of the control rodent [9, 10].

To replace the damaged hippocampal circuitry and rehabilitate the impaired memory, Berger and his colleagues developed an implantable biomimetic computational model that can act as an artificial bridge between CA3 and CA1 regions [9, 10] (Fig. 5.1b, bottom). The biomimetic device consists of three components: a multi-electrode array (MEA) recording the neural activities from the spared CA3 region, a prediction model of spiking activity of CA1 neurons based on the spiking activity of CA3 neurons, called multi-input/multi-output (MIMO) nonlinear model [39] implemented in a very large scale integration (VLSI) microchip [40], and a stimulating MEA that can stimulate CA1 region with the MIMO-predicted neural activities. In this paradigm, the neural activities recorded from the MEA in the CA3 region are fed into the MIMO model so that it can predict the spatio-temporal patterns of CA1 neuronal spikes, which are directly delivered to the damaged CA1 region through the stimulating MEA.

The key to the biomimetic chip is the development of the MIMO model. The MIMO model was developed to predict the spatio-temporal spike patterns of post-synaptic neurons from the pre-synaptic activity patterns using a Volterra series approach [41]. In the MIMO model, the spiking activity of neural population was predicted from the presynaptic input spike train based on a five electrophysiological basis: (i) the feedforward process transforming the input spike train to synaptic potential, (ii) the feedback process generating an after-potential caused by the output spike, (iii) the intrinsic neuronal noise causing neural variability, (iv) subthreshold potential dynamics, and (v) the threshold function to generate the output spike. The model dynamics were computed by the equations:

$$\omega = u(k, x) + a(h, y) + \varepsilon(\sigma)$$

$$y = \begin{cases} 0 & when\ \omega < \theta \\ 1 & when\ \omega \geq \theta \end{cases}$$

where x is the input spike trains, y is the output spike, θ is the threshold for output spike, k is the feedforward kernel for transforming input, h is the feedback kernel for modulating after-potential, σ is the noise variable, and u, a, ε, and ω are the resultant continuous hidden variables representing dynamics for feedforward transformation, feedback transformation, noise, and summed subthreshold potential, respectively [41]. To adjust the parameters of the MIMO model, spike trains of CA3 and CA1 region during the DNMS task in the control rodents were recorded [39, 41].

When the developed artificial bridge containing the MEA and MIMO system was directly tested in the damaged CA3-CA1 connection in the lesioned rodent, the rodent was able to successfully perform the DNMS task [9]. The success rate of the DNMS task was increased up to 90 % compared to that of the control rodent, demonstrating that the artificial bridge has a potential as the memory prosthetic device.

Recently, the application and configuration of the MIMO model have been widely expanded to non-human primates [36, 42], confirming the possibility of the MIMO model as a human memory prosthetic device. Using the same DNMS task, the implanted biomimetic electrical chip containing the MIMO model successfully recovered the impaired memory function of the non-human primate [36, 42]. Moreover, the trained memory could even be transplanted to untrained animal using the MIMO model [43]. These results suggested that the MIMO model could act as a storage system for hippocampal memory, further underscoring the roles of the MIMO model as the computational basis of memory prosthetics [44]. Therefore, implementation of the MIMO model on a VLSI hardware microchip [40] could be used for a real-time portable memory prosthesis for the memory implantation.

5.2.3 Implementation of Neuromorphic Network Model to MIMO

The limitation of the conventional DBS is that it is an open-loop system which generates only the pre-programmed electrical stimulations without being able to receive any feedback from the current brain state. Therefore, a closed-loop biomimetic electric chip that can simultaneously record neural activity and stimulate the brain based on the current state of the brain activity is advantageous to the open-loop paradigm for neural prosthetic devices [45–47]. To generate the appropriate stimulation parameter, the closed-loop system needs to acquire real-time signals from the brain and decode the acquired neural signals properly *ad-hoc*, quickly switching between the two modes of operation. Thus, the development of a closed-loop feedback chip implementing the MIMO model is needed but with the current technology, designing such closed-loop systems itself is a major engineering challenge.

Another limitation of the current MIMO model is that it is a hard-wired model. However, real synapses are not hard-wired but it displays changes in synaptic efficacy during experience [17], such as LTP and LTD. If synaptic plasticity is implemented to the MIMO model, the input-output dynamics of the MIMO model could be fundamentally altered. Therefore, the memory implantation using the stationary MIMO model could be applicable in only simple memory task on a well-trained animal. To tackle this problem, recent studies have developed synaptic plasticity-modulated prediction of output activity approximated by Volterra series [48, 49] in the MIMO model [44]. However, the synaptic plasticity implemented in the recent MIMO model [44] is too simple to capture the entirety of the complex circuit and molecular dynamics occurring during synaptic plasticity *in vivo* [50–52]. For example, learning rule is dependent on the types [53–55] and the locations of synapses [53]. In addition, there are more than 40 types of neuron in the hippocampus [56, 57], and their intricate interactions during learning process [58] hinder the development of a simple prediction model of learning process. For the

last decades, biophysically detailed computational models have been able to capture the complex dynamics of learning process [59–62]. Ionic and molecular detailed synaptic plasticity model could capture the complex molecular dynamics of synaptic plasticity [59, 60] and detailed hippocampal circuitry model containing various types of neurons showed complex dynamics of learning process [61, 62]. Therefore, implementation of the neuromorphic neural network model with biophysically detailed synaptic learning rule onto the VLSI hardware [63] could provide the possibility to integrate the biophysically detailed hippocampal learning model onto the closed-loop memory prosthetic device which would accelerate the progress in memory implantation (Fig. 5.1b).

5.3 Conclusion

Implantation of closed-loop feedback neuromorphic chips that can simultaneously record and stimulate the hippocampal neurons could be used not only to replace the damaged hippocampal circuit but also to enhance or implant memory in the future. Although a promising future technique, the possibility of hippocampal damage due to the implantation of memory prosthesis itself is another major challenge remained for the medical application. A better way to manipulate the brain would be to develop a non-invasive technique that can activate neural populations. Recently, it has been shown that optogenetic activation of only a small population of hippocampal neurons could modulate memory [64–66]. In addition, transcranial direct current stimulation technique that delivers focal electrical stimulation to the cortical surface has proven to enhance long-term visual memory in Alzheimer's patients [67]. Therefore, with a better understanding of the neuronal mechanisms underlying memory function, non-invasive brain stimulation techniques combined with the closed-loop neuromorphic chip technology could be the way forward to develop true memory implantation technology in the future. If successful, developing memory prosthesis that can implant or remove specific memories from a human brain may provide fundamental treatment methodologies for patients with memory dysfunction such as Alzheimer's disease and dementia.

References

1. Masters CL, Simms G, Weinman NA, Multhaup G, McDonald BL, Beyreuther K (1985) Amyloid plaque core protein in Alzheimer disease and Down syndrome. Proc Natl Acad Sci U S A 82:4245–4249
2. Selkoe DJ (1994) Cell biology of the amyloid beta-protein precursor and the mechanism of Alzheimer's disease. Annu Rev Cell Biol 10:373–403
3. Alzheimer's A (2014) 2014 Alzheimer's disease facts and figures. Alzheimers Dement 10:e47–e92

4. Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M (2010) Alzheimer's disease: clinical trials and drug development. Lancet Neurol 9:702–716
5. Halgren E, Wilson CL (1985) Recall deficits produced by afterdischarges in the human hippocampal formation and amygdala. Electroencephalogr Clin Neurophysiol 61:375–380
6. Halgren E, Wilson CL, Stapleton JM (1985) Human medial temporal-lobe stimulation disrupts both formation and retrieval of recent memories. Brain Cogn 4:287–295
7. Coleshill SG, Binnie CD, Morris RG, Alarcon G, van Emde Boas W, Velis DN et al (2004) Material-specific recognition memory deficits elicited by unilateral hippocampal electrical stimulation. J Neurosci 24:1612–1616
8. Berger TW, Gerhardt G, Liker MA, Soussou W (2008) The impact of neurotechnology on rehabilitation. IEEE Rev Biomed Eng 1:157–197
9. Berger TW, Hampson RE, Song D, Goonawardena A, Marmarelis VZ, Deadwyler SA (2011) A cortical neural prosthesis for restoring and enhancing memory. J Neural Eng 8:046017
10. Hampson RE, Song D, Chan RH, Sweatt AJ, Riley MR, Goonawardena AV et al (2012) Closing the loop for memory prosthesis: detecting the role of hippocampal neural ensembles using nonlinear models. IEEE Trans Neural Syst Rehabil Eng 20:510–525
11. Zola-Morgan S, Squire LR, Amaral DG (1986) Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. J Neurosci 6:2950–2967
12. Andersen P, Bliss TV, Skrede KK (1971) Lamellar organization of hippocampal pathways. Exp Brain Res 13:222–238
13. Rempel-Clower NL, Zola SM, Squire LR, Amaral DG (1996) Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci 16:5233–5255
14. Milner B, Taylor L, Sperry RW (1968) Lateralized suppression of dichotically presented digits after commissural section in man. Science 161:184–186
15. Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20:11–21
16. Bliss TV, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361:31–39
17. Bliss TV, Lomo T (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 232:331–356
18. Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275:213–215
19. Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472
20. Ito M, Kano M (1982) Long-lasting depression of parallel fiber-Purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. Neurosci Lett 33:253–258
21. Dudek SM, Bear MF (1992) Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. Proc Natl Acad Sci U S A 89:4363–4367
22. Bienenstock EL, Cooper LN, Munro PW (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J Neurosci 2:32–48
23. Wixted JT, Squire LR, Jang Y, Papesh MH, Goldinger SD, Kuhn JR et al (2014) Sparse and distributed coding of episodic memory in neurons of the human hippocampus. Proc Natl Acad Sci U S A 111:9621–9626
24. Davis KD, Taub E, Houle S, Lang AE, Dostrovsky JO, Tasker RR et al (1997) Globus pallidus stimulation activates the cortical motor system during alleviation of Parkinsonian symptoms. Nat Med 3:671–674
25. Limousin P, Krack P, Pollak P, Benazzouz A, Ardouin C, Hoffmann D et al (1998) Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. N Engl J Med 339:1105–1111

26. Williams JM, Givens B (2003) Stimulation-induced reset of hippocampal theta in the freely performing rat. Hippocampus 13:109–116
27. Ehret A, Haaf A, Jeltsch H, Heimrich B, Feuerstein TJ, Jackisch R (2001) Modulation of electrically evoked acetylcholine release in cultured rat septal neurones. J Neurochem 76:555–564
28. Penfield W, Perot P (1963) The brain's record of auditory and visual experience. A final summary and discussion. Brain 86:595–696
29. Stone SS, Teixeira CM, Devito LM, Zaslavsky K, Josselyn SA, Lozano AM et al (2011) Stimulation of entorhinal cortex promotes adult neurogenesis and facilitates spatial memory. J Neurosci 31:13469–13484
30. Toda H, Hamani C, Fawcett AP, Hutchison WD, Lozano AM (2008) The regulation of adult rodent hippocampal neurogenesis by deep brain stimulation. J Neurosurg 108:132–138
31. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C et al (2005) Deep brain stimulation for treatment-resistant depression. Neuron 45:651–660
32. Lozano AM, Mayberg HS, Giacobbe P, Hamani C, Craddock RC, Kennedy SH (2008) Subcallosal cingulate gyrus deep brain stimulation for treatment-resistant depression. Biol Psychiatry 64:461–467
33. Suthana N, Haneef Z, Stern J, Mukamel R, Behnke E, Knowlton B et al (2012) Memory enhancement and deep-brain stimulation of the entorhinal area. N Engl J Med 366:502–510
34. Hardenacke K, Shubina E, Buhrle CP, Zapf A, Lenartz D, Klosterkotter J et al (2013) Deep brain stimulation as a tool for improving cognitive functioning in Alzheimer's dementia: a systematic review. Front Psychiatry 4:159
35. Suthana N, Fried I (2014) Deep brain stimulation for enhancement of learning and memory. Neuroimage 85(3):996–1002
36. Hampson RE, Song D, Opris I, Santos LM, Shin DC, Gerhardt GA et al (2013) Facilitation of memory encoding in primate hippocampus by a neuroprosthesis that promotes task-specific neural firing. J Neural Eng 10:066013
37. Gaffan D (1974) Recognition impaired and association intact in the memory of monkeys after transection of the fornix. J Comp Physiol Psychol 86:1100–1109
38. Mishkin M (1978) Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. Nature 273:297–298
39. Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW (2009) Nonlinear modeling of neural population dynamics for hippocampal prostheses. Neural Netw 22:1340–1351
40. Berger TW, Song D, Chan RH, Marmarelis VZ, LaCoss J, Wills J et al (2012) A hippocampal cognitive prosthesis: multi-input, multi-output nonlinear modeling and VLSI implementation. IEEE Trans Neural Syst Rehabil Eng 20:198–211
41. Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW (2007) Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses. IEEE Trans Biomed Eng 54:1053–1066
42. Hampson RE, Gerhardt GA, Marmarelis V, Song D, Opris I, Santos L et al (2012) Facilitation and restoration of cognitive function in primate prefrontal cortex by a neuroprosthesis that utilizes minicolumn-specific neural firing. J Neural Eng 9:056012
43. Deadwyler SA, Berger TW, Sweatt AJ, Song D, Chan RH, Opris I et al (2013) Donor/recipient enhancement of memory in rat hippocampus. Front Syst Neurosci 7:120
44. Song D, Harway M, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW (2014) Extraction and restoration of hippocampal spatial memories with non-linear dynamical modeling. Front Syst Neurosci 8:97
45. Rosin B, Slovik M, Mitelman R, Rivlin-Etzion M, Haber SN, Israel Z et al (2011) Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. Neuron 72:370–384
46. Santos FJ, Costa RM, Tecuapetla F (2011) Stimulation on demand: closing the loop on deep brain stimulation. Neuron 72:197–198
47. Jarosiewicz B, Masse NY, Bacher D, Cash SS, Eskandar E, Friehs G et al (2013) Advantages of closed-loop calibration in intracortical brain-computer interfaces for people with tetraplegia. J Neural Eng 10:046012

48. Robinson BS, Song D, Berger TW (2013) Laguerre-Volterra identification of spike-timing-dependent plasticity from spiking activity: a simulation study. Conf Proc IEEE Eng Med Biol Soc 2013:5578–5581
49. Dong S, Robinson BS, Granacki JJ, Berger TW (2014) Implementing spiking neuron model and spike-timing-dependent plasticity with generalized Laguerre-Volterra models. Conf Proc IEEE Eng Med Biol Soc 2014:714–717
50. Sjostrom PJ, Turrigiano GG, Nelson SB (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32:1149–1164
51. Froemke RC, Dan Y (2002) Spike-timing-dependent synaptic modification induced by natural spike trains. Nature 416:433–438
52. Wang HX, Gerkin RC, Nauen DW, Bi GQ (2005) Coactivation and timing-dependent integration of synaptic potentiation and depression. Nat Neurosci 8:187–193
53. Froemke RC, Poo MM, Dan Y (2005) Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature 434:221–225
54. Woodin MA, Ganguly K, Poo MM (2003) Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. Neuron 39:807–820
55. Lamsa KP, Kullmann DM, Woodin MA (2010) Spike-timing dependent plasticity in inhibitory circuits. Front Synaptic Neurosci 2:8
56. Somogyi P, Klausberger T (2005) Defined types of cortical interneurone structure space and spike timing in the hippocampus. J Physiol 562:9–26
57. Klausberger T, Somogyi P (2008) Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. Science 321:53–57
58. Saraga F, Balena T, Wolansky T, Dickson CT, Woodin MA (2008) Inhibitory synaptic plasticity regulates pyramidal neuron spiking in the rodent hippocampus. Neuroscience 155:64–75
59. Rubin JE, Gerkin RC, Bi GQ, Chow CC (2005) Calcium time course as a signal for spike-timing-dependent plasticity. J Neurophysiol 93:2600–2613
60. Urakubo H, Honda M, Froemke RC, Kuroda S (2008) Requirement of an allosteric kinetics of NMDA receptors for spike timing-dependent plasticity. J Neurosci 28:3310–3323
61. Hasselmo ME, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci 15:5249–5262
62. Cutsuridis V, Cobb S, Graham BP (2010) Encoding and retrieval in a model of the hippocampal CA1 microcircuit. Hippocampus 20:423–446
63. Rachmuth G, Shouval HZ, Bear MF, Poon CS (2011) A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity. Proc Natl Acad Sci U S A 108:E1266–E1274
64. Liu X, Ramirez S, Pang PT, Puryear CB, Govindarajan A, Deisseroth K et al (2012) Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature 484:381–385
65. Ramirez S, Liu X, Lin PA, Suh J, Pignatelli M, Redondo RL et al (2013) Creating a false memory in the hippocampus. Science 341:387–391
66. Liu X, Ramirez S, Tonegawa S (2014) Inception of a false memory by optogenetic manipulation of a hippocampal memory engram. Philos Trans R Soc Lond B Biol Sci 369:20130142
67. Boggio PS, Ferrucci R, Mameli F, Martins D, Martins O, Vergari M et al (2012) Prolonged visual memory enhancement after direct current stimulation in Alzheimer's disease. Brain Stimul 5:223–230

Chapter 6
Moving Brain-Controlled Devices Outside the Lab: Principles and Applications

Robert Leeb*, Ricardo Chavarriaga*, Serafeim Perdikis, Iñaki Iturrate, and José d.R. Millán

Abstract This chapter provides an overview of the functionality and the underlying principles of the brain-computer interfaces (BCI) developed by the Chair in Non-Invasive Brain-Machine Interface (CNBI) of the Swiss Federal Institute of Technology (EPFL), as well as exemplary applications where those have been successfully evaluated. Our laboratory mainly develops *non-invasive* BCI systems based on electroencephalographic (EEG) signals and, thus, devoid of medical hazards, real-time, portable, relatively low-cost and minimally obtrusive. Our research is pushing forward *asynchronous* paradigms offering a spontaneous, user-driven and largely ecological interaction. Furthermore, we stand on the *machine-learning* way to BCI with emphasis on personalization, configurability and adaptability, coupled with *mutual learning* training protocols, so that elaborate signal processing and pattern recognition methods are optimally combined with the user's learnable modulation of brain signals towards high and robust performances and universal usability. Additionally, *cognitive mental state* monitoring is employed to shape or refine the interaction. *Shared-control* approaches allow smart, context-aware robotics to complement the BCI channel for more fine-grained control and reduction of the user's mental workload. Last but not least, *hybrid BCI* designs exploit additional physiological signals to augment the BCI modality and enrich the control paradigm, thus also exploiting potential residual capabilities of disabled end-users.

The applicability and effectiveness of the aforementioned principles is hereby demonstrated in four exemplary applications evaluated with both able-bodied and

*These authors contributed equally to this work.

R. Leeb
Center for Neuroprosthetics, École Polytechnique Fédérale de Lausanne, Station 19, CH-1015 Lausanne, Switzerland
e-mail: robert.leeb@epfl.ch

R. Chavarriaga • S. Perdikis • I. Iturrate • J.d.R. Millán (✉)
Defitech Chair in Brain-Machine Interface, École Polytechnique Fédérale de Lausanne, Station 11, CH-1015 Lausanne, Switzerland
e-mail: ricardo.chavarriaga@epfl.ch; serafeim.perdikis@epfl.ch; inaki.iturrate@epfl.ch; jose.millan@epfl.ch

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_6

motor-disabled end-users. These applications include a hybrid, motor imagery (MI)-based speller, a telepresence robot equipped with shared-control, cognitive mental state monitoring paradigms able to recognize and correct errors, and, finally, a car driving application where a passive BCI enabled on a smart car assists towards increased safety and improved driving experience. Remarkably, our results show that the performance of end-users with disabilities was similar to that of a group of healthy users, who were more familiar with the experiment and the environment. This demonstrates that end-users are able to successfully use BCI technology.

Keywords Brain-computer interface • Shared control • Hybrid control • Motor imagery • Error potentials • Cognitive states • Spelling • Telepresence robot • Car driving

6.1 Introduction

In the course of the last thirty years, Brain-Computer Interface (BCI) technology has been gradually establishing itself as the solution towards the dream of direct mind control by human individuals [88]. Since the first demonstration of its profound applicability [6], BCI research has showcased rapid development and reached several milestones. Consequently, BCI prototypes are nowadays increasingly deployed in real-world conditions, outside the comfort and manageability of the laboratory [8, 44, 61].

Within the general scope of augmenting human abilities with novel, brain-actuated control channels, clinical applications have always been the spearhead of BCI [50]. People suffering from severe motor disabilities and, especially, locked-in individuals have the most to gain out of a non-muscular control channel [6] and naturally comprise the primary user group of BCI. Such end-users have been shown to benefit from a variety of BCI prototypes, targeting, among others, communication [32, 41, 66], assistive mobility [10, 25, 45] motor substitution [17, 31] and restoration [73], environmental control [2, 26] and, more recently, neurorehabilitation [3, 72, 79].

In spite of the field's current relative maturity, several obstacles still hinder the critical transition of BCI technology from the lab to an everyday use in people's homes or in clinics. To begin with, the available neural interfaces and brain imaging methods currently employed for BCI are associated with at least a number of limitations [63] including invasiveness (and, hence, medical risks and lack of long-term stability), low time or spatial resolution and brain tissue coverage, compromised signal-to-noise ratios (SNR), signal quality degradation, reduced portability, obtrusiveness and, certainly, high monetary cost. Furthermore, BCI training protocols can be lengthy, cumbersome, unintuitive and unable to exploit the full scale of the brain's immense learning and plasticity capacities, essentially failing to bring a large percentage of prospective users into control of a BCI [49, 85]. Most importantly, BCI control is known to still be fairly crude, suffer low information transfer rates (ITR) and exhibit intense performance fluctuations

which call for frequent system calibrations [87]. The latter require constant expert supervision, raising the issue of the systems' degree of robustness, independence and automaticity [44]. Additionally, many BCI paradigms lack the ability to provide spontaneous, ecological interaction. In the clinical setting, problems arise regarding the coordination of the multi-faceted effort needed (involving professionals of various expertise) and the process of incorporating professional and end-user feedback into BCI system design [44, 92].

Pre-clinical and real-world validation of state-of-the-art prototypes in the widest possible application spectrum are essential for identifying principles that can, on one hand, alleviate the aforementioned deficiencies to the largest extent possible and, on the other hand, allow to maximize the added value of BCI applications even with an imperfect BCI channel at hand. This chapter is devoted to providing an overview of such successfully applied principles by the Chair in Non-Invasive Brain-Machine Interface (CNBI) of the Swiss Federal Institute of Technology (EPFL), as well as exemplary applications where those have been evaluated.

In summary, our laboratory mainly develops *non-invasive* BCI systems based on electroencephalographic (EEG) signals and, thus, devoid of medical hazards, real-time, portable, relatively low-cost and minimally obtrusive. Our research is pushing forward *asynchronous* paradigms offering a spontaneous, user-driven and largely ecological interaction. Furthermore, we stand on the *machine-learning* way to BCI with emphasis on personalization, configurability and adaptability, coupled with *mutual learning* training protocols, so that elaborate signal processing and pattern recognition methods are optimally combined with the user's learnable modulation of brain signals towards high and robust performances and universal usability. Additionally, *cognitive mental state* monitoring is employed to shape or refine the interaction. *Shared-control* approaches allow smart, context-aware robotics to complement the BCI channel for more fine-grained control and reduction of the user's mental workload. Last but not least, *hybrid BCI* designs exploit additional physiological signals to augment the BCI modality and enrich the control paradigm, thus also exploiting potential residual capabilities of disabled end-users.

The applicability and effectiveness of the aforementioned principles is hereby demonstrated in four exemplary applications evaluated with both able-bodied and motor-disabled end-users. These include a hybrid, motor imagery (MI)-based speller [66], a telepresence robot [45, 81] equipped with shared-control, cognitive mental state monitoring paradigms able to recognize and correct errors [34], and, finally, a car driving application where a passive BCI enabled on a smart car assists towards increased safety and improved driving experience [15].

6.2 Principles for Brain Control

Before presenting applications of brain-controlled devices, we will first explain the underlying principles and give a short introduction to the neurophysiological basis of the brain correlates we exploit for interaction.

6.2.1 *Brain-Computer Interaction: It Takes Two to Tango*

For a human user to convey mental commands to a device, we have first to record their brain activity, from which the BCI infers the user's intention. In our case, we rely on brain electrical signals (EEG in particular) because of their high temporal resolution, which allows real-time operation of devices. EEG is also a non-invasive, cheap, secure and practical method [52].

Several types of BCIs exist and various control methods can be applied. In our direct control examples presented later on (see Sect. 6.3), we utilize a BCI based on motor imagery, and so record the EEG over the central and sensorimotor cortex. MI is described as the mental rehearsal of a motor act without any overt motor output [18], which activates similar brain regions to those engaged during planning and preparation of such real movements [20, 35]. The imagination of different types of movements (e.g., right hand, left hand or feet), results in an amplitude suppression (known as event-related desynchronization, ERD [68]) or in an amplitude enhancement (event-related synchronization, ERS) of Rolandic mu rhythm (7–13 Hz) and the central beta rhythm (13–30 Hz) recorded over the sensorimotor cortex of the participant [69].

Figure 6.1a illustrates and describes the basic elements of a BCI. In essence, a BCI distinguishes different patterns of brain activity, each being associated to a particular intention or mental task. Hence, adaptation is a key component of a

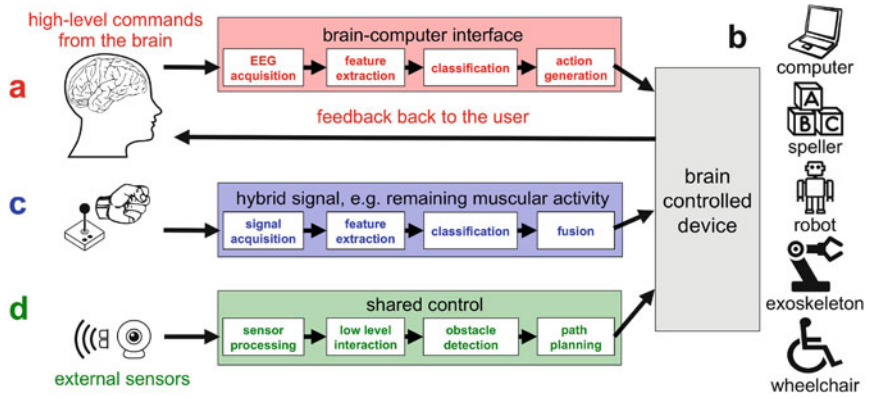

Fig. 6.1 (**a**, *red*) Basic principle of a BCI: The electrical signals from the brain are acquired and discriminant, stable features are extracted. These are then classified to generate actions to control devices. The participant receives immediate feedback from the output of the BCI and/or the generated action of the device. (**b**, *black*) Examples of brain-controlled devices. (**c**, *blue*) Hybrid principle: In parallel to the BCI also other control signals (like muscular activity or input devices like joysticks, switches) can be used for interaction in a hybrid BCI. (**d**, *green*) The shared control principle: Besides the user issuing high-level commands via the BCI (usually at a lower pace than the operation of the device), the system is acquiring fast and precise the environmental information (via sonars, webcams . . .). The shared control system combines the two sources of information to better estimate the user's intent and improve task execution

BCI, because, on the one side, users must learn to modulate their neural activity so as to generate distinct brain patterns, while, on the other side, machine learning techniques ought to discover the individual brain patterns characterizing the mental tasks executed by the user. In essence, a BCI is a two-learner system that must engage in a mutual adaptation process [12, 58, 86]. In our case, this translates into selecting discriminant, stable features—namely, user-specific brain components that maximize the separability between mental commands [24, 55] and that, because of the non-stationary nature of brain signals, are stable over time—to build optimal models to decode the user's intention. These initial features represent those brain components that the user can naturally modulate and, via feedback received during online BCI training, learn to control quickly and voluntarily.

Based on the inherent properties of the EEG, the number of commands that can be detected is small. Consequently, the possible information transfer rate of a BCI is quite limited, compared to our normal healthy way of interaction. More details about the experimental paradigm, signal processing, machine learning techniques (feature extraction, feature selection, classification and evidence accumulation), training principle, and feedback are given in [44].

In addition, the interaction with the brain-controlled devices can also exploit brain activity elicited by the feedback perceived by the user. Specifically, a BCI system can decode the brain responses evoked by the system's actions and use this information to infer cognitive processes taking place during the interaction. One example of this case is the decoding of error-related brain activity. These are neural responses that appear when the user is aware of erroneous decisions made by the device [13, 22]. The possibility of knowing when the user perceives BCI errors can be then used to correct or improve the system's performance [16]. An example of such applications is presented in Sect. 6.3.3.

6.2.2 Shared Control: Making Possible the Impossible

Interaction with real devices or complex computer programs via a BCI can be quite challenging. For example, driving a wheelchair or a robot in a natural environment (scattered with obstacles like chairs, tables, people . . .) demands a fine and quickly responding control signal like a joystick. Unfortunately, given the current state-of-the-art BCIs, users rarely attain this fine grade of control—at least not continuously. Nevertheless, despite the mismatch between task requirements and users' skills, researchers have successfully demonstrated the feasibility of mentally controlling complex robotic devices from EEG.

A key factor to do so is the use of smart interaction designs, which in the field of robotics corresponds to shared control [9, 56, 84]. In our example, a smart wheelchair will help the user to navigate through the crowded environment. The user issues via the BCI the high level commands such as left, right and forward, which are then interpreted by the wheelchair controller based on the contextual information from its sensors (obstacles perceived) and the status of the device itself

(position and velocities) to better estimate the user's intent (see Fig. 6.1b). Based on these interpretations, the wheelchair can perform intelligent maneuvers (e.g., obstacle avoidance, guided turnings).

Context awareness or shared control is a key component of any future BCI system, as it will shape the closed-loop dynamics between the user and the brain-actuated device so that tasks can be performed as easily as possible and effectively, but not autonomously. A critical aspect for BCI is coherent feedback—the behavior of the robotic device should be intuitive to the user and the robot should unambiguously understand the user's mental commands. Otherwise, people find it difficult to form mental models of neuroprostheses, such as a brain-controlled exoskeleton, that extend their natural bodies.

6.2.3 Hybrid Control: Explore All Possibilities

Practical brain-computer interfaces for disabled people should allow them to use all their remaining functionalities, so as to truly enhance their capabilities. Sometimes these people have residual muscular activity, most likely in the morning when they are not exhausted. In the framework of a user-center approach, all existing and reliable control signals should be exploited. In such a hybrid approach, we could combine and enhance conventional assistive products (operated using some residual muscular functionality) by BCI technology, leading to what is called a hybrid BCI (see Fig. 6.1c).

As a general definition, a hybrid BCI is a combination of different input signals including at least one BCI channel [59, 70]. Thus, it could be a combination of multiple brain signals: such as MI for control and detection of error-related potentials for correction of false commands [23], MI with SSVEP [71] or P300 [1], SSVEP with P300 [89], and even two brain modalities such as NIRS and EEG [21].

More importantly, hybrid BCIs truly augment end-users' capabilities by combining a BCI with their residual biosignals (such as muscular signals, etc.) or special assistive technology input devices (e.g., joysticks, switches, etc.). There exist a few examples of combining brain and other biosignals: switching a standard SSVEP BCI on/off via a heart rate variation [77], or fusing electromyographic (EMG) with EEG activity [42] so that the subjects could achieve a good control of their hybrid BCI independently of their level of muscular fatigue. Millán et al. [59] review the state of the art and challenges in combining hybrid BCI and assistive technologies.

6.3 Applications

In this section we describe examples of BCI controlled applications for motor-disabled users and healthy participants, in the area of 'Communication & Control', 'Motor Substitution', 'Error detection', and 'Mental & Cognitive State Monitoring'.

6.3.1 *Hybrid BCI Spelling for End-Users*

Among BCI applications, text-entry has been the best-studied field, as restoring communication skills is the highest priority of end-users.[1] The main limitation remains the lack of clinical evaluation of spelling prototypes, currently largely limited to P300-based BCI spellers (e.g., [41]). The latter require subjects to focus their attention on flashing stimuli and a positive waveform occurring approximately 300 ms after the infrequent task-relevant stimulus is detected. Furthermore, great progress has been achieved in utilizing asynchronous spontaneous approaches for spelling prototypes. Recently, our group focused on the clinical evaluation of a novel motor imagery-based, hybrid BCI spelling prototype, called *BrainTree*, by 6 disabled end-users and 10 able-bodied individuals. Following a user-centered approach, the BrainTree speller is the first MI-based BCI system to be validated by a fairly large pool of end-users [66]. We are thus providing a competitive BCI alternative to P300 spelling, which, like most BCI paradigms, cannot be successfully employed by all prospective end-users [28], while also being unable to support self-paced control.

Figure 6.2a illustrates the speller's graphical user interface (GUI), where characters are arranged alphabetically. A vertical red cursor, the "caret", denotes the current position, while an orange "bubble" surrounds the characters currently available. Underneath the character bar, the user observes a conventional MI BCI feedback. The user employs one of two MI tasks to move the caret towards the desired character. The procedure is repeated until the latter is the only one left within the "bubble", in which case it will be typed after the next transition and a

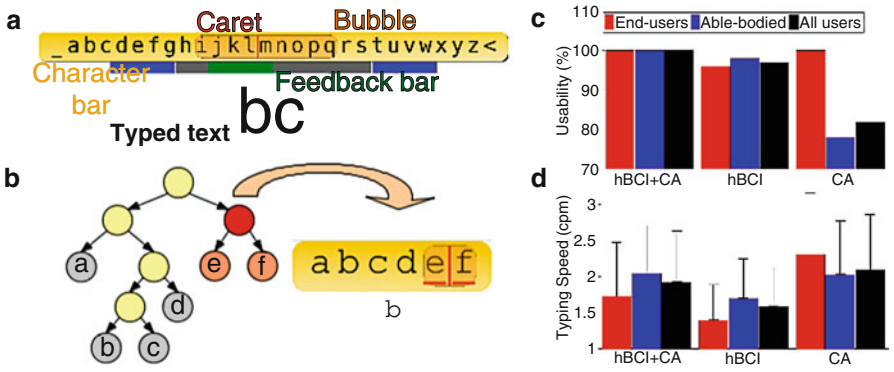

Fig. 6.2 BrainTree speller. (**a**) Graphical user interface. (**b**) Example of underlying binary tree structure. (**c**) Usability as flawless task completion rate per condition (%). (**d**) Average typing speed as characters per minute (cpm) per condition

[1]This Section follows closely a prior published paper by the authors [66].

new typing round is initiated. This simple GUI hides the underlying complexity, where characters are the leaf nodes of a binary tree. A simple example on a reduced dictionary is illustrated in Fig. 6.2b. Therein, the "caret" represents the tree's current internal node, the "bubble" surrounds the leaf node characters within the current node's two subtrees and a BCI command moves to the left/right child node. Effectively, in each typing round, each character is associated with a binary "codeword" of left/right transitions.

The BrainTree prototype largely owes its effectiveness and efficiency to the incorporation of the shared-control and hybrid BCI principles described in Sect. 6.2.2. Shared-control increases the spelling efficiency in BrainTree by exploiting the Hu-Tucker entropy coding algorithm [33], which reduces the average length of character codewords to the theoretical minimum, while also preserving alphabetic ordering. Entropy coding is made possible through a trained prefix-based language model. From the user perspective, shared-control refines the user's control by deciding the optimal ending position of the "caret". Besides the main BCI control modality, a secondary control modality (hybrid approach) for "undo" commands is implemented through monitoring electromyographic activity for the detection of a user's brisk movement, in case residual muscular activity allows it. Equivalently, any other residual ability (e.g., eye-blinks, sip-and-puff) can be used. The "undo" functionality enables immediate correction of an erroneous transition, thus being more efficient than the, also available, "backspace" functionality ("pure"-BCI approach).

The BrainTree user evaluation study comprises two phases, the BCI training sessions described in [44], and the spelling sessions, performed in real-world conditions involving noisy and crowded environments (clinics, lab, end-user's home). The latter consist of four copy-spelling tasks, repeated across three conditions. The first condition, termed *hBCI+CA*, where both the hybrid BCI (hBCI) mode and context-awareness (CA) through shared-control are enabled, is meant to evaluate the overall usability and efficiency of BrainTree. The *hBCI* (shared-control disabled) and *CA* (hybrid mode disabled) conditions are introduced to experimentally quantify the added value of shared-control and hybrid error-handling. A total of 6 end-users (all male, 39 $\pm$ 15.5), and 10 able-bodied users (9 male and 1 female, 29.0 $\pm$ 4.6) participated in the final BrainTree evaluation. For logistic reasons only 2 end-users performed the *CA* condition.

The most impressive outcome is the speller's impeccable usability for both user categories, as measured through flawless (without time restrictions) spelling task completion rates and shown in Fig. 6.2c. Usability is only seriously compromised in *CA*, revealing the significant added value a hybrid modality might bring. Still, the speller is largely usable in a "pure"-BCI fashion (*CA*, 82 % task completion for all users), while also both end-users trying this version completed all tasks successfully. The positive impact of the implemented shared-control approach is showcased in Fig. 6.2d, which reports the average typing speed per condition as the number of characters written per minute (cpm). Subjects achieved a 17 % typing speed reduction (from 1.92 cpm in *hBCI+CA*, to 1.59 cpm in *hBCI* across all users) on average when disabling the optimal entropy coding. End-users underperformed for

hBCI+CA compared to able-bodied ones in terms of spelling speed by 0.3 cpm on average. This is not the case for usability, however, showing the clinical applicability of the BrainTree prototype. These figures are only slightly inferior to the ones reported for P300 spelling.

In summary, we have presented a first MI BCI spelling, end-user evaluation study, proving the extensive usability and competitive efficiency of our speller [66]. Furthermore, we have demonstrated the added value of hybrid BCI designs and shared-control approaches. Finally, we argue that our user-centered approach can be accredited with the successful incorporation of such principles into the speller's design, underlining the importance of pushing BCI technology out of the lab.

6.3.2 *End-Users Out There: Controlling a Telepresence Robot*

Another BCI application area in which disabled people will benefit in the future is motor substitution and assisted mobility.[2] Thereby, disabled end-users will be able to remotely control a telepresence robot, their own wheelchair or an exoskeleton directly via the BCI. Since the timing and speed of interaction is crucial for the interaction with these neuroprosthetic devices, shared control will help the user to achieve such a complex and time demanding task by using smart devices, which incorporate information of the environment (see Sect. 6.2.2). Although the whole field of neuroprosthetics targets disabled people with motor impairments as end-users, all successful demonstrations of brain-controlled robots or neuroprosthetics, except [17, 44, 62], have been actually carried out with either healthy human subjects or non-human primates.

Here, we describe our recent work [45], during which nine end-users with severe motor disabilities, were mentally driving a telepresence robot in a natural office environment from their clinic up to 550 km away and compare their performances to a set of ten healthy users carrying out the same tasks in our lab. Remarkably, the system functioned effectively although the patients had never visited the location where the telepresence robot was operating. For controlling, the user asynchronously sent high-level commands for turning to the left or right (with the help of a motor-imagery based BCI) to achieve the desired goals, while short-term, low-level interaction for obstacle avoidance was done by the shared control (see Fig. 6.1d and Sect. 6.2.2). In the applied shared control paradigm, the default behavior of the robot is to move forward at a constant speed, and to pro-actively slow down and turn to avoid obstacles as it approaches them. The robot is based on Robotino™ by FESTO (Esslingen, Germany), a small circular mobile platform, which is equipped with nine infrared sensors that can detect obstacles up to ∼30 cm distance (see Fig. 6.3a). Furthermore, a notebook with a camera is added on top

[2]This Section follows closely a prior published paper by the authors [45].

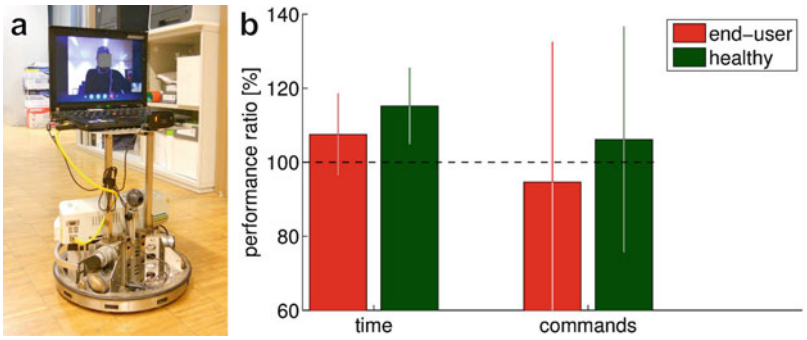

Fig. 6.3 (**a**) The telepresence robot equipped with nine circular distributed infrared sensors for obstacle detection and a laptop with the Skype™connection on top. (**b**) Performance ratio between the time and the number of commands required to complete the task when using BCI versus when using the manual input device. The grand average over nine end-users with motor disabilities and ten healthy participants is plotted together with the standard deviation

of the robot for telepresence purposes, so that the participant can interact with the remote environment via Skype™.

The experiment was performed twice, once with BCI control with the help of shared control, and once, as a reference condition, with direct manual control via keyboard buttons. In case of the end-users with motor disabilities, some of them were able to press buttons on a modified keyboard and others were using head switches. Generally, we used whatever residual motion they were capable of producing. The subject's task was to bring the robot to four predefined target positions within the natural working space. As performance measures, we measured the time and number of commands needed to reach the targets.

The most striking result of the experiment is that all end-users succeeded to mentally control the telepresence robot as efficiently as the healthy participants, who were familiar with the environment. The mean ratio of time to complete the task for the BCI condition compared with the manual condition is only $109.2 \pm 11.1\,\%$ for the end-users, as compared to $115.1 \pm 10.3\,\%$ for the healthy participants. Shared control also helped all subjects (including naive BCI subjects or users with disabilities) to complete a rather complex task in similar time and with similar number of commands to those required by manual commands without shared control (see Fig. 6.3b). More details are given in [45].

Based on these results, we argue that end-users with motor disabilities can mentally control a telepresence robot via a BCI as good as healthy participants. Furthermore, the proposed shared control approach reduces the subjects' cognitive workload, as it assists them in coping with low-level navigation issues (e.g., obstacle avoidance) and allows the subject to focus the attention on his final destination. Thereby shared control helps BCI users to maintain attention for longer periods of time (since the amount of BCI commands can be reduced and their precise timing is not so critical).

6.3.3 *Error-Related Potentials*

A successful BCI application should not only have the ability to decode and execute high-level commands, but also to recover itself from errors committed by the user or even the machine.[3] Interestingly, there have been numerous works demonstrating the existence of distinct neural imprints associated to user's own detection of errors [60]. These correlates, the error-related potentials or ErrP, are also evoked when the user observes another human or a machine committing an error [16]. Any efficient and robust BCI application should also target the decoding of these signals, so as to recover from wrongly executed commands or use them to adapt the system [7, 22, 76]. In this section, we show the existence of these signals under different experimental protocols, their similarity across experiments and demonstrate the feasibility of their single-trials decoding.

Twelve participants performed three experimental protocols of increasing complexity as shown in Fig. 6.4a. They were seated on a comfortable chair facing the visual displays of the protocols approximately one meter away, and instructed to restrict eye movements and blinks to specific resting periods. In all experiments they were asked to evaluate whether a device moves towards a given target location or not. The device moved in discrete steps and the time between movements was randomly chosen within the range 1.7–4.0 s. There was a probability of moving in the wrong direction of about 30 %. Experiments were always performed in the same order from the simplest to the most complex one. The first experiment, E_1, consists of a cursor that moves in discrete steps (either left or right) towards a target. In the second protocol, E_2, the user monitors a simulated robotic arm that moves on a 2D plane (allowed movement directions were left, right, up and down). The third experiment, E_3, consists of the same task using a real robotic arm. A detailed

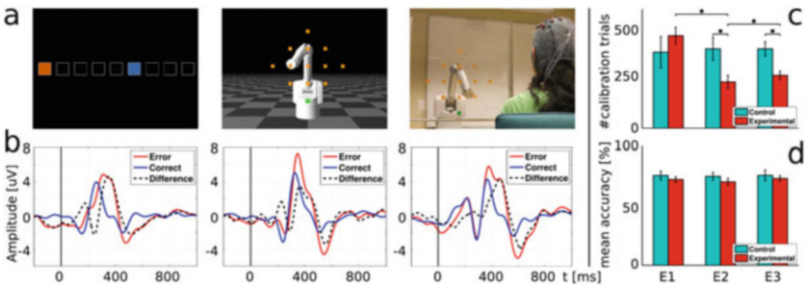

Fig. 6.4 Error-related potentials. (**a**) Experimental protocols: (*Left*) One-dimensional cursor movement (E_1). (*Center*) Two- dimensional movements of a simulated robotic arm (E_2). (*Right*) Real robotic arm (E_3). (**b**) ErrPs at the FCz electrode obtained in each experimental protocol. (**c**) Number of calibration trials needed on each experiment for the control group (*blue*) and the experimental group (*red*) (*$p < 0.05$). (**d**) Mean online accuracy for each experiment and group

[3]This Section follows closely a prior published paper by the authors [34].

explanation of the protocols and methods is provided in [34]. Each experiment started by a calibration phase. This phase had a variable length depending on the obtained performance. Calibration stopped whenever the mean accuracy (ten-fold cross validation) on the training data exceeds 75 %. Then the classifier parameters were fixed, and performance was tested on an online phase lasting 400 trials. For the classification, features from eight fronto-central channels were selected in the window of 200 to 800 ms using a spatiotemporal filter. Single-trials were classified as erroneous or correct using linear discriminant analysis (LDA).

Figure 6.4b shows the averaged ErrP of each experiment with respect to the action onset (0 ms). The error potential can be easily characterized by the difference grand average (error minus correct, in dashed), with two early positive and negative peaks in fronto-central sites, followed by two larger positive and negative peaks. Despite no differences were found on the peak amplitudes of the ErrPs (ANOVA tests, $p = 0.510$ and $p = 0.391$ respectively), the experimental protocol affected the latency of the ErrP elicitation ($p < 0.01$). Interestingly, these variations can be easily estimated by computing the cross-correlation between the grand-average ErrPs for each experiment. In brief, given a previous experiment E_i, from which data is available, and a new experiment E_j, the ErrP latency variation $d_{E_iE_j}$ will correspond to the shift that yields the maximum cross-correlation. Then, data from E_i can be shifted in time by $d_{E_iE_j}$ and used along the available (few) trials from the new experiment E_j to train a classifier. We thus compared whether this latency correction mechanism effectively reduces the calibration time. To that end, we defined two groups of participants depending on the training procedure. The control group (N = 6) followed a standard calibration approach, i.e., based only on data from the current experiment. The experimental group (N = 6) used latency-corrected trials from the previous experiment to build the classifier for the current task. That is, standard calibration was followed for E_1, while data from that experiment was used during the calibration period of E_2. Similarly, during calibration for E_3, the data from E_2, was used. The latency between experiments was estimated based on the cross-correlation of the difference ErrP (error minus correct condition, see Fig. 6.4b).

Figure 6.4c shows the number of calibration trials needed in each experiment to reach the stopping criteria. The calibration period for the control group was similar for all experiments. In contrast, the experimental group exhibits a large reduction on the required calibration trials in E_2 and E_3 when previous information was re-used. The ANOVA test revealed a significant interaction between the experiment and group ($p = 0.002$). On the other hand, no significant difference was found in the decoding performance (Fig. 6.4d) for all experiments and subjects ($p > 0.85$). These results indicate that, provided data from previous experiments, knowledge from these protocols can be transferred to the new task using the latency correction algorithm.

These results confirmed the existence of error potentials under several experimental protocols, and how they generalize across different tasks of increasing complexity. Compensating for latency variations between protocols has proven to shorten the calibration phase in new applications without affecting the decoding performance, higher than 75 % on average. Their existence and generalization across

different tasks will boost their applicability for BCI applications. Indeed, future brain-computer interfaces will exploit these signals to create more sophisticated systems able to avoid wrong behaviors or even learn and adapt to the user's preferences.

6.3.4 BCI Decoding of Driver's Cognitive States

Previous sections describe BCI applications focused on substitution or restoration of communication and motor capabilities of individuals with severe disabilities.[4] These have been traditionally the main target population of BCI systems. However, applications for able-bodied persons have also been proposed, in particular in very specific contexts like space exploration (e.g. tackling situational disabilities) [19], military applications (e.g. target recognition from satellite images) [75], or games [40, 51, 53, 80]. Along these lines, we study the potential application of BCI technology to improve the interaction with intelligent cars.

Current cars are able to provide assistance for different driving maneuvers—from parking and lane changing assistance to fully autonomous navigation. We propose that BCIs can be used to monitor the driver's cognitive state and provide this information to modulate the assistance provided by the intelligent car. In this case, the driver is kept in the loop, allowing the driving support system to take into account the external context (as perceived by in-car sensors) and the user's intention (as decoded from electroencephalographic signals) to provide suitable and timely assistance; seamlessly interacting with the driver.

Previous works have focused on decoding driver's level of attention and emergency braking from brain activity [29, 30, 38, 39, 48]. Extending these, we studied EEG correlates of cognitive states in order to predict future actions or to evaluate whether the decisions of the intelligent system are coherent with the user intentions. Preliminary works have already shown the feasibility of decoding anticipation and error-related EEG potentials in realistic virtual environments and during human-robot interaction [14, 67]. We thus focused on assessing the decoding of these types of signals while driving. Following the principles of shared control (c.f. Sect. 6.2.2) information about the driver's cognitive state (i.e. internal context), combined with external environmental cues captured by in-car sensors allows the evaluation of the instantaneous needs of the user [11]. Imagine a case where the car approaches an intersection, sensors in the car can perceive a red traffic light while brain signals can tell whether the driver is aware of it and is getting ready to stop [37]. This information, appearing as early as 500 ms before any noticeable action or muscular activity, allows for the assistance to be provided only if the driver was unprepared to act.

[4]This Section follows closely prior published papers by the authors [27, 37].

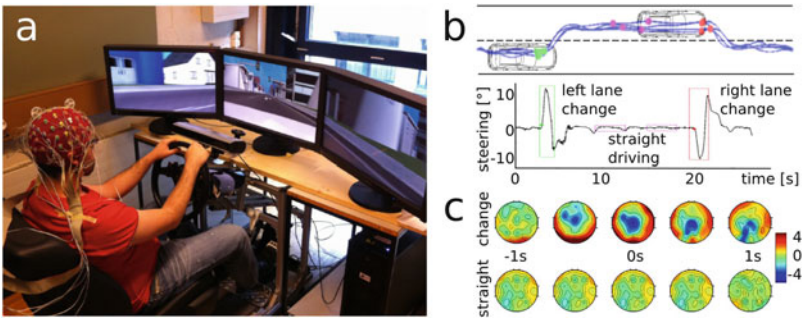

Fig. 6.5 (**a**) Experimental setup: Driving simulator used in the experiments, including a real car-seat and steering wheel, and three 3D monitors. (**b**) *Top*. Car trajectories during lane changes in a simulated highway. Blue traces correspond to the different car paths aligned to the time when the lane change is performed (t = 0). Dots represent different behaviors as indicated by the steering angle (*bottom*). *Green*: lane change towards the left; *Red*: lane change towards the right; *Magenta*: Driving in a straight trajectory. *Bottom*. Steering pattern obtained from the driving simulator logger. (**c**) Grand average ERP during lane changes (*top*) and straight driving periods (*bottom*). Topographical activity represented by a top view of the scalp (nose up). t = 0 corresponds to the moment of steering [27]

We have performed experiments in a realistic car-driving simulator where we can record simultaneously the car steering angles; brake and acceleration pedals, with EEG, EOG (electrooculogram) and EMG signals (see Fig. 6.5a). Using an experimental protocol based on the scenario described above (i.e. approaching a traffic light) we have identified EEG correlates of anticipation and movement preparation highly consistent with those observed in classical experimental paradigms. Single-trial decoding of these correlates can thus be used as a marker of the user's preparedness to execute the expected action [37]. Classification of slow-cortical potentials (0.1–1 Hz) using a quadratic discriminant analysis (QDA) classifier shows high performance in the decoding of anticipation-related potentials preceding braking or accelerating actions (18 subjects; 4-fold cross-validation; average AUC equal to 0.79 and 0.83, respectively). In another experiment where subjects had to drive in a simulated highway, we observed slow negative EEG deflections over central areas—akin to the motion related potential—appearing more than 1 s prior to self-paced lane changes [27], as shown in Fig. 6.5c. Classification of these potentials yields a true positive rate of 68.8 ± 6.6 (5-fold cross-validation), with average detection times of 641 ± 94 ms before the actual steering action (N = 6).

As mentioned above, intelligent cars can use information from these potentials, as well as environmental information and previous experience to provide assistance to the driver. Nonetheless, these devices are prone to errors that may hinder their performance. As discussed in Sect. 6.3.3, BCI systems can exploit as well error-related brain signals [13, 16]. We therefore studied whether these signals could also be observed and decoded when a driver assistant provides information to the user. In the experimental protocol, the system was presenting the driver its

prediction about upcoming turning directions. Coincident with previous studies in other experimental protocols, a stereotypic response is elicited in fronto-central areas when the feedback provided by the car does not match the user's intention. Experiments with 22 subjects show clear statistical differences between error and correct conditions between 200 and 600 ms after the feedback. More importantly, error-related activity was recognized above chance level (average AUC across subjects = 0.70; 10-fold cross-validation) [90].

To summarize, we consistently found neural signatures of anticipation, movement preparation and error processing elicited during simulated driving. Despite the fact that EEG signals in this realistic scenario are more susceptible to be affected by environmental artifacts, these signals can be successfully decoded in single trials. Therefore providing information about cognitive processes, taking place while driving intelligent cars, is feasible. We are currently assessing this framework in closed loop settings as well as in a real car setting to fully evaluate its potential. In our approach, information about internal user states, in combination with contextual information gathered by in-car sensors is exploited to better tune the human-machine interaction. In consequence, the intelligent car—instead of superseding the driver—will provide timely and tailored assistance only when required; thus keeping the user in the loop and allowing him to fully enjoy the pleasures of driving.

6.4 Summary, Discussion and Outlook

In this chapter we gave an overview of the functionality and the underlying principles of our brain-computer interface and showed different scenarios and applications. We emphasized communication and control applications for motor-disabled end-users, like spelling on a computer and controlling of a telepresence robot. Furthermore, we presented applications in which healthy participants will benefit from a BCI, such as error processing, anticipation, and movement preparation.

The introduced principles are of key importance for the use of BCI system outside the laboratory environments, in real applications. The selection of prototypes described in this chapter illustrates how the principles of (i) shared control and context awareness, as well as (ii) hybridness can help in overcoming some of the BCI limitations and enable users to achieve complex tasks. Remarkably, our results show that the performance of end-users with disabilities was similar to that of a group of healthy users, who were more familiar with the experiment and the environment. This demonstrates that end-users are able to successfully use BCI technology. However, BCIs are not yet ready for an independent use at home [44] and some gaps for usability and reliability have to be addressed.

The benefits of shared control are outstanding in the telepresence experiment, where all subjects completed the navigation task in similar time and with similar number of commands to those required by manual commands without context awareness. Thus, we argue that context awareness reduces the subjects' cognitive workload as it: (i) assists them in coping with low-level navigation issues (such

as obstacle avoidance and allows the subject to focus the attention on his final destination) and (ii) helps BCI users to keep attention for longer periods of time (since the amount of BCI commands can be reduced and their precise timing is not so critical).

The hybrid BCI approach allows the use of all remaining functionalities of an end-user and enables more degree of freedom in control applications. A hybrid BCI can exploit different input signals either independently (each controlling a different output dimension), alternatively (switching from one to another depending on their reliability, which may change due to fatigue), or in parallel (via fusion). In Sect. 6.3.1 we showed that muscular activity can be used to correct errors of a motor-imagery based BCI [66], which could be also done via error potential [22, 23]. Results show how this approach boosted performance, even for users with a relatively low BCI performance. In another work we demonstrated that multimodal fusion techniques allow the combination of brain control with other residual motor control signals and thereby achieve better and more reliable performances, independently of the level of muscular fatigue [42].

The exploitation of cognitive signals, which carry valuable information about the users' state in parallel to the detection of the mental commands they want to deliver, is pushing BCIs to a new level of robustness, flexibility and performance. Information like how users perceive and value the interaction, their level of attention, or their awareness to erroneous decisions, will allow us to modulate the level of support needed at any time for effective BCI control [74]. We can envision a future where systems (e.g., an intelligent car) would be able to modify their behavior depending on the user's cognitive state so as to provide variable levels of assistance, and even become more proactive in case the user's attention to the task decreases. Similarly, detection of cognitive states would enable to extend the principle of adaptive shared control to rehabilitation scenarios as the system will continuously adjust the level of assistance as the patient recovers his/her motor capabilities.

Although not discussed in this chapter, BCI technology is used in motor rehabilitation, especially in case of stroke patients. The underlying idea is that BCIs provide a mean to access the damaged motor network of the brain after stroke, and could be used to drive and promote beneficial plasticity, especially in the early rehabilitation stages, when no motor output can be generated. The BCI decodes the attempt to execute a hand movement and, if successful, provides visual feedback on a screen or activates either a robot or functional electrical stimulation (FES) in order to actually execute the desired movement of the affected limb. Several studies have shown the positive rehabilitation effect of such a therapeutic approach [3, 4, 72] and even showed signs of promoting cortical reorganization even during chronic phase.

As mentioned above, progress is still needed in order to translate BCI systems into practical systems that can be widely deployed and used independently by their intended users. Although we expect great advancement in the next years because of the attention that the BCI field is receiving from funding agencies and companies, there are still some challenges ahead. The first one is to exploit other kinds of information contained in brain signals recorded from multiple areas beyond the usual time-frequency analysis, as well as additional correlates of voluntary behavior

so as to improve the performance of the BCI and achieve natural, effortless operation of complex devices. In the former case, we can analyze the dynamics of brain activity through the use of features based on EEG connectivity patterns [5, 91]. In the latter case, we can decode not only the type of imagined movement, but also the onset of those movements [46], directions of the intended movements [47], and where in the surrounding space the user is covertly attending [82, 83].

The second challenge is to design physical interfaces that can operate permanently and, ideally, last a lifetime. New hardware spans from dry EEG electrodes to biocompatible and fully implantable neural interfaces including ECoG, LFP, and SUA from multiple brain areas. An essential component of all of them is wireless transmission and ultra low-power consumption. This new hardware for life-long recordings generates other challenges. One of them is the need for continuous adaptation of the decoders to the evolving brain signals of the subject that work in a transparent way while the user operates the brain-controlled device [54, 57, 65, 85]. A second one is to find efficient ways of combining practical hybrid BCI tools with smart interaction designs and devices to facilitate use over long periods of time and to reduce the cognitive load.

Another future area for BCI is to extract cognitive-relevant information in order to improve standard interactions, which is becoming increasingly interesting for healthy users. A further component that will facilitate intuitive and natural control of BCIs is the incorporation of rich multimodal feedback to convey information to the user about the decisions made by the device and its internal states. This feedback can involve several sensory modalities [36, 43, 78] or direct stimulation of the brain [64].

Attempts to address each, or subsets, of these challenges and future research avenues are already being tackled. Yet, large and long-term user studies are missing and are essential for BCI to become a mature technology. In this respect, translational studies involving end-users operating brain-controlled devices in their natural environments (home, public transport, office, or clinical settings) will identify additional challenges. Only adopting a user-centered iterative approach (for the end-users as well as for the healthy population) will allow addressing the specific needs and requirements of the different future user groups.

Acknowledgements The authors acknowledge the use of text from their prior publications [27, 34, 37, 45, 66] in Sect. 6.3.

This work was supported by the European ICT programme projects TOBI: Tools for Brain-Computer Interaction (FP7-224631) and Opportunity: Activity and Context Recognition with Opportunistic Sensor Configuration (ICT-225938), the Swiss National Science Foundation NCCR Robotics, Nissan Motor Co. Ltd., the Hasler Foundation, and the dissemination by the European ICT coordination and support action BNCI Horizon 2020 (FP7-ICT-609593). The writing up of this work was also supported by the Swiss canton of Valais-Wallis and by the City of Sion. This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

References

1. Allison BZ, Brunner C, Kaiser V, Müller-Putz GR, Neuper C, Pfurtscheller G (2010) Toward a hybrid brain–computer interface based on imagined movement and visual attention. J Neural Eng 7:026007
2. Aloise F, Schettini F, Aricó P, Leotta F, Salinari S, Mattia D, Babiloni F, Cincotti F (2011) P300-based brain-computer interface for environmental control: an asynchronous approach. J Neural Eng 8:025025
3. Ang KK, Guan C, Chua KS, Ang BT, Kuah C, Wang C, Phua KS, Chin ZY, Zhang H (2011) A large clinical study on the ability of stroke patients to use an EEG-based motor imagery brain-computer interfaces. Clin EEG Neurosci 42(4):253–258
4. Biasiucci A, Leeb R, Al-Khodairy A, Buhlmann V, Millán JdR (2013) Motor recovery after stroke by means of BCI-guided functional electrical stimulation. In: Proceedings of 5th international BCI meeting, Asilomar, California, US
5. Billinger M, Brunner C, Müller-Putz GR (2013) Single-trial connectivity estimation for classification of motor imagery data. J Neural Eng 10(4):046006
6. Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H (1999) A spelling device for the paralysed. Nature 398(6725):297–298
7. Blankertz B, Dornhege G, Schäfer C, Krepki R, Kohlmorgen J, Müller K-R, Kunzmann V, Losch F, Curio G (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans Neural Syst Rehabil Eng 11:127–131
8. Blankertz B, Tangermann M, Vidaurre C, Fazli S, Sannelli C, Haufe S, Maeder C, Ramsey L, Sturm I, Curio G et al (2010) The Berlin brain–computer interface: non-medical uses of BCI technology. Front Neurosci 4:198
9. Carlson T, Demiris Y (2008) Human-wheelchair collaboration through prediction of intention and adaptive assistance. In: Proceedings of IEEE international conference on robotics and automation, Pasadena, pp 3926–3931
10. Carlson T, Millán JdR (2013) Brain–controlled wheelchairs: a robotic architecture. IEEE Robot Autom Mag 20(1):65–73
11. Carlson T, Leeb R, Chavarriaga R, Millán JdR (2012) Online modulation of the level of assistance in shared control systems. In: Proceedings of IEEE international conference on systems man and cybernetics, Seoul, pp 3321–3326
12. Carmena JM (2013) Advances in neuroprosthetic learning and control. PLoS Biol 11(5):e1001561
13. Chavarriaga R, Millán JdR (2010) Learning from EEG error-related potentials in noninvasive brain-computer interfaces. IEEE Trans Neural Syst Rehabil Eng 18(4):381–388
14. Chavarriaga R, Perrin X, Siegwart R, Millán JdR (2012) Anticipation- and error-related EEG signals during realistic human-machine interaction: a study on visual and tactile feedback. In: Proceedings of 34th annual international conference of the IEEE engineering in medicine and biology society, San Diego, pp 6723–6726
15. Chavarriaga R, Gheorghe L, Zhang H, Khaliliardali Z, Millán JdR (2013) Detecting cognitive states for enhancing driving experience. In: Proceedings of 5th international brain-computer interface meeting 2013, Pacific Grove, Asilomar, California, US
16. Chavarriaga R, Sobolewski A, Millán JdR (2014) Errare machinale est: the use of error-related potentials in brain-machine interfaces. Front Neurosci 8:208
17. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, McMorland AJC, Velliste M, Boninger ML, Schwartz AB (2013) High-performance neuroprosthetic control by an individual with tetraplegia. Lancet 381(9866):557–564
18. Decety J (1996) The neurophysiological basis of motor imagery. Behav Brain Res 77(1–2):45–52
19. de Negueruela C, Broschart M, Menon C, Millán JdR (2010) Brain-computer interfaces for space applications. J Pers Ubiquitous Comput 15(5):527–537

20. Ehrsson HH, Geyer S, Naito E (2003) Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. J Neurophysiol 90:3304–3316
21. Fazli S, Mehnert J, Steinbrink J, Curio G, Villringer A, Müller K-R, Blankertz B (2012) Enhanced performance by a hybrid NIRS–EEG brain computer interface. Neuroimage 59:519–529
22. Ferrez PW, Millán JdR (2008) Error-related EEG potentials generated during simulated brain-computer interaction. IEEE Trans Biomed Eng 55:923–929
23. Ferrez PW, Millán JdR (2008) Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy. In: Proceedings of 4th international brain-computer interface workshop and training course, Graz, pp 197–202
24. Galán F, Ferrez PW, Oliva F, Guàrdia J, Millán JdR (2007) Feature extraction for multi-class BCI using canonical variates analysis. In: Proceedings of IEEE international symposium on intelligent signal processing, Alcalá de Henares
25. Galán F, Nuttin M, Lew E, Ferrez PW, Vanacker G, Philips J, Millán JdR (2008) A brain-actuated wheelchair: asynchronous and non-invasive brain-computer interfaces for continuous control of robots. Clin Neurophysiol 119(9):2159–2169
26. Gao X, Xu D, Cheng M, Gao S (2003) A BCI-based environmental controller for the motion-disabled. IEEE Trans Neural Syst Rehabil Eng 11:137–140
27. Gheorghe LA, Chavarriaga R, Millán JdR (2013) Steering timing prediction in a driving simulator task. In: Proceedings of 35th annual international conference of the IEEE engineering in medicine and biology society, Osaka, Japan, pp 6913–6916
28. Guger C, Edlinger G, Harkam W, Niedermayer I, Pfurtscheller G (2003) How many people are able to operate an EEG-based brain-computer interface (BCI)? IEEE Trans Neural Syst Rehabil Eng 11:145–147
29. Haufe S, Treder MS, Gugler MF, Sagebaum M, Curio G, Blankertz B (2011) EEG potentials predict upcoming emergency brakings during simulated driving. J Neural Eng 8(5):056001
30. Haufe S, Kim J-W, Kim I-H, Sonnleitner A, Schrauf M, Curio G, Blankertz B (2014) Electrophysiology-based detection of emergency braking intention in real-world driving. J Neural Eng 11:056011
31. Hochberg LR, Bacher D, Jarosiewicz B, Masse NY, Simeral JD, Vogel J, Haddadin S, Liu J, Cash SS, van der Smagt P, Donoghue JP (2012) Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature 485(7398):372–375
32. Höhne J, Holz E, Staiger-Sälzer P, Müller K-R, Kübler A, Tangermann M (2014) Motor imagery for severely motor-impaired patients: evidence for brain-computer interfacing as superior control solution. PLoS ONE 9:e104854
33. Hu TC, Tucker AC (1971) Optimal computer search trees and variable-length alphabetical codes. SIAM J Appl Math 21(4):514–532
34. Iturrate I, Chavarriaga R, Montesano L, Minguez J, Millán JdR (2014) Latency correction of error potentials between different experiments reduces calibration time for single-trial classification. J Neural Eng 11:036005
35. Jeannerod M, Frak V (1999) Mental imaging of motor activity in humans. Curr Opin Neurobiol 9(6):735–739
36. Kaufmann T, Holz EM, Kübler A (2013) Comparison of tactile, auditory and visual modality for brain-computer interface use: a case study with a patient in the locked-in state. Front Neurosci 7:129
37. Khaliliardali Z, Chavarriaga R, Gheorghe LA, Millán JdR (2012) Detection of anticipatory brain potentials during car driving. In: Proceedings of 34th annual international conference of the IEEE engineering in medicine and biology society, San Diego, vol 2012, pp 3829–3832
38. Kim I-H, Kim J-W, Haufe S, Lee S-W (2015) Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. J Neural Eng 12:016001

39. Kohlmorgen J, Dornhege G, Braun ML, Blankertz B, Müller KR, Curio G, Hagemann K, Bruns A, Schrauf M, Kincses WE (2007) Improving human performance in a real operating environment through real-time mental workload detection. In: Dornhege G, Millán JdR, Hinterberger T, McFarland D, Müller KR (eds) Toward brain-computer interfacing. MIT Press, Cambridge, pp 409–422
40. Krepki R, Blankertz B, Curio G, Müller K-R (2007) The Berlin brain-computer interface (BBCI): towards a new communication channel for online control in gaming applications. J Multimed Tools Appl 33:73–90
41. Kübler A, Birbaumer N (2008) Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol 119(11):2658–2666
42. Leeb R, Sagha H, Chavarriaga R, Millán JdR (2011) A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. J Neural Eng 8(2):025011
43. Leeb R, Gwak K, Kim D-S, Millán JdR (2013) Freeing the visual channel by exploiting vibrotactile BCI feedback. In: Proceedings of 35th annual international conference of the IEEE engineering in medicine and biology society, Osaka, Japan
44. Leeb R, Perdikis S, Tonin L, Biasiucci A, Tavella M, Molina A, Al-Khodairy A, Carlson T, Millán JdR (2013) Transferring brain-computer interfaces beyond the laboratory: successful application control for motor-disabled users. Artif Intell Med 59(2):121–132
45. Leeb R, Tonin L, Carlson T, Millán JdR (2015) Towards independence: a BCI telepresence robot for people with severe motor disabilities. Proc IEEE 103:969–982
46. Lew E, Chavarriaga R, Silvoni S, Millán JdR (2012) Detection of self-paced reaching movement intention from EEG signal. Front Neuroeng 5:13
47. Lew E, Chavarriaga R, Silvoni S, Millán JdR (2014) Single trial prediction of self-paced reaching directions from EEG signals. Front Neurosci 8:222
48. Lin C-T, Ko L-W, Sheng T-K (2009) Computational intelligent brain computer interaction and its applications on driving cognition. IEEE Comput Intell Mag 4(4):32–46
49. Lotte F, Larrue F, Mühl C (2013) Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. Front Hum Neurosci 7:568
50. Mak JN, Wolpaw EW (2009) Clinical applications of brain-computer interfaces: current state and future prospects. IEEE Rev Biomed Eng 2:187–199
51. Marshall D, Coyle D, Wilson S, Callaghan M (2013) Games, gameplay, and BCI: the state of the art. IEEE Trans Comput Intell AI Games 5(2):82–99
52. Mason SG, Bashashati A, Fatourechi M, Navarro KF, Birch GE (2007) A comprehensive survey of brain interface technology designs. Ann Biomed Eng 35(2):137–169
53. Millán JdR (2003) Adaptive brain interfaces. Commun ACM 46:74–80
54. Millán JdR (2004) On the need for on-line learning in brain-computer interfaces. In: Proceedings of international joint conference on neural networks, Budapest
55. Millán JdR, Franzé M, Mouriño J, Cincotti F, Babiloni F (2002) Relevant EEG features for the classification of spontaneous motor-related tasks. Biol Cybern 86:89–95
56. Millán JdR, Renkens F, Mouriño J, Gerstner W (2004) Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans Biomed Eng 51(6):1026–1033
57. Millán JdR, Buttfield A, Vidaurre C, Krauledat M, Schögl A, Shenoy P, Blankertz B, Rao RPN, Cabeza R, Pfurtscheller G, Müller K-R (2007) Adaptation in brain-computer interfaces. In: Dornhege G, Millán JdR, Hinterberger T, McFarland D, Müller KR (eds) Toward brain-computer interfacing. MIT Press, Cambridge, pp 303–325
58. Millán JdR, Ferrez PW, Galán F, Lew E, Chavarriaga R (2008) Non-invasive brain-machine interaction. Int J Pattern Recognit Artif Intell 22(5):959–972
59. Millán JdR, Rupp R, Müller-Putz GR, Murray-Smith R, Giugliemma C, Tangermann M, Vidaurre C, Cincotti F, Kübler A, Leeb R, Neuper C, Müller KR, Mattia D (2010) Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. Front Neurosci 4:161

60. Miltner WHR, Braun CH, Coles MGH (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. J Cogn Neurosci 9(6):788–798
61. Müller K-R, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J Neurosci Methods 167:82–90
62. Müller-Putz GR, Scherer R, Pfurtscheller G, Rupp R (2005) EEG-based neuroprosthesis control: a step towards clinical practice. Neurosci Lett 382:169–174
63. Navarro X, Krueger TB, Lago N, Micera S, Stieglitz T, Dario P (2005) A critical review of interfaces with the peripheral nervous system for the control of neuroprostheses and hybrid bionic systems. J Peripher Nerv Syst 10(3):229–258
64. O’Doherty JE, Lebedev MA, Ifft PJ, Zhuang KZ, Shokur S, Bleuler H, Nicolelis MAL (2011) Active tactile exploration using a brain-machine-brain interface. Nature 479:228–231
65. Orsborn AL, Moorman HG, Overduin SA, Shanechi MM, Dimitrov DF, Carmena JM (2014) Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. Neuron 82(6):1380–1393
66. Perdikis S, Leeb R, Williamson J, Ramsay A, Tavella M, Desideri L, Hoogerwerf E-J, Al-Khodairy A, Murray-Smith R, Millán JdR (2014) Clinical evaluation of BrainTree, a motor imagery hybrid BCI speller. J Neural Eng 11(3):036003
67. Perrin X, Chavarriaga R, Colas F, Siegwart R, Millán JdR (2010) Brain-coupled interaction for semi-autonomous navigation of an assistive robot. Robot Auton Syst 58(12):1246–1255
68. Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 110:1842–1857
69. Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain-computer communication. Proc IEEE 89:1123–1134
70. Pfurtscheller G, Allison BZ, Bauernfeind G, Brunner C, Solis Escalante T, Scherer R, Zander TO, Müller-Putz G, Neuper C, Birbaumer N (2010) The hybrid BCI. Front Neurosci 4:42
71. Pfurtscheller G, Solis-Escalante T, Ortner R, Linortner P, Muller-Putz GR (2010) Self-paced operation of an SSVEP-based orthosis with and without an imagery-based “brain switch”: a feasibility study towards a hybrid BCI. IEEE Trans Neural Syst Rehabil Eng 18:409–414
72. Ramos-Murguialday A, Broetz D, Rea M, Läer L, Yilmaz Ö, Brasil FL, Liberati G, Curado MR, Garcia-Cossio E, Vyziotis A, Cho W, Agostini M, Soares E, Soekadar S, Caria A, Cohen LG, Birbaumer N (2013) Brain-machine interface in chronic stroke rehabilitation: a controlled study. Ann Neurol 74(1):100–108
73. Rohm M, Schneiders M, Müller C, Kreilinger A, Kaiser V, Müller-Putz GR, Rupp R (2013) Hybrid brain-computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury. Artif Intell Med 59(2):133–142
74. Saeedi S, Carlson T, Chavarriaga R, Millán JdR (2013) Making the most of context-awareness in brain-computer interfaces. In: Proceedings of IEEE international conference on cybernetics, Lausanne, Switzerland
75. Sajda P, Pohlmeyer E, Wang J, Parra LC, Christoforou C, Dmochowski J, Hanna B, Bahlmann C, Singh MK, Chang S-F (2010) In a blink of an eye and a switch of a transistor: cortically coupled computer vision. Proc IEEE 98(3):462–478
76. Schalk G, Wolpaw JR, McFarland DJ, Pfurtscheller G (2000) EEG-based communication: presence of an error potential. Clin Neurophysiol 111:2138–2144
77. Scherer R, Müller-Putz GR, Pfurtscheller G (2007) Self-initiation of EEG-based brain-computer communication using the heart rate response. J Neural Eng 4:L23–L29
78. Schreuder M, Blankertz B, Tangermann M (2010) A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLoS ONE 5(4):e9813
79. Silvoni S, Ramos-Murguialday A, Cavinato M, Volpato C, Cisotto G, Turolla A, Piccione F, Birbaumer N (2011) Brain-computer interface in stroke: a review of progress. Clin EEG Neurosci 42(4):245–252
80. Tangermann MW, Krauledat M, Grzeska K, Sagebaum M, Vidaurre C, Blankertz B, Müller K-R (2008) Playing pinball with non-invasive BCI. In: Advances in neural information processing systems 21, Vancouver, pp 1641–1648

81. Tonin L, Leeb R, Tavella M, Perdikis S, Millán JdR (2010) The role of shared-control in BCI-based telepresence. In: Proceedings of 2010 IEEE international conference on systems, man and cybernetics, Istanbul, pp 1462–1466
82. Tonin L, Leeb R, Millán JdR (2012) Time-dependent approach for single trial classification of covert visuospatial attention. J Neural Eng 9(4):045011
83. Tonin L, Leeb R, Sobolewski A, Millán JdR (2013) An online EEG BCI based on covert visuospatial attention in absence of exogenous stimulation. J Neural Eng 10(5):056007
84. Vanhooydonck D, Demeester E, Nuttin M, Van Brussel H (2003) Shared control for intelligent wheelchairs: an implicit estimation of the user intention. In: Proceedings of 1st international workshop advances in service robot, Bardolino, pp 176–182
85. Vidaurre C, Kawanabe M, von Bünau P, Blankertz B, Müller K-R (2011) Toward unsupervised adaptation of LDA for brain-computer interfaces. IEEE Trans Biomed Eng 58(3):587–597
86. Vidaurre C, Sannelli C, Müller K-R, Blankertz B (2011) Co-adaptive calibration to improve bci efficiency. J Neural Eng 8:025009
87. Wolpaw J, Winter-Wolpaw E (2012) Brain-computer interfaces: principles and practice. Oxford University Press, Oxford/New York
88. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain-computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
89. Yin E, Zhou A, Jiang J, Chen F, Liu Y, Hu D (2013) A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. J Neural Eng 10:026012
90. Zhang H, Chavarriaga R, Gheorghe LA, Millán JdR (2013) Inferring driver's turning direction through detection of error related brain activity. In: Proceedings of 35th annual international conference of the IEEE engineering in medicine and biology society, Osaka, vol 2013, pp 2196–2199
91. Zhang H, Chavarriaga R, Millán JdR (2014) Towards implementation of motor imagery using brain connectivity features. In: Proceedings of 6th international brain-computer interface conference, Graz, Austria
92. Zickler C, Di Donna V, Kaiser V, Al-Khodairy A, Kleih S, Kuebler A, Malavasi M, Mattia D, Mongardi S, Neuper C, Rohm M, Rupp R (2009) Brain computer interaction applications for people with disabilities: defining user needs and user requirements. In: Proceedings of AAATE, Florence, p 5

Part III
Big Data Neurocomputing

Chapter 7
Across Cultures: A Cognitive and Computational Analysis of Emotional and Conversational Facial Expressions in Germany and Korea

Christian Wallraven, Dong-Cheol Hur, and Ahyoung Shin

Abstract Humans use a wide variety of communicative signals – among those, facial expressions play a key role in communicating not only emotional, but also more general, non-verbal signals. Here, we present results from a combined cognitive and computational analysis of emotional and conversational facial expressions in the context of cross-cultural research. Using two large databases of dynamic facial expressions, we show that both Western and Asian observers structure the interpretation space of a large range of facial expressions using the same two evaluative dimensions (valence and arousal). In addition, several computational experiments show the advantage of using graph-models for automatic recognition of facial expressions, since these models are able to capture the complex dynamics and inter-dependence of the movements of facial features in the face.

Keywords Facial expressions • Cross-cultural psychology • Emotions • Conversational expressions • Graph models

7.1 Introduction

Human communication can be divided into verbal and non-verbal signals. In the case of non-verbal signals, the human face plays a key role: the face itself conveys the person's identity and additional kinds of attributes such as attractiveness, intelligence, and trustworthiness, for example. Importantly, when the face starts to move, facial expressions are produced that convey information about one's feelings,

C. Wallraven (✉) • A. Shin
Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea
e-mail: wallraven@korea.ac.kr

D.-C. Hur
Department of Computer Science and Engineering, Korea University, Seoul 136-713, Republic of Korea

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*, Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_7

emotions, or intentions – the face starts to communicate. This "language of facial expressions" allows for rich and efficient interaction between people and forms one of the most important parts of non-verbal communication. It is therefore not surprising that there is a considerable amount of research about facial expressions in various fields, focusing on both cognitive aspects (that is, the investigation of how humans perceive, process, and use facial expressions) and computational aspects (that is, the investigation of how one may teach computers to understand and react to human facial expressions).

First, concerning cognitive aspects, an interesting debate in the field has been whether certain kinds of facial expressions may be "universal" signals of communication. For example, some cross-cultural studies have found evidence for highly robust interpretation and recognition of a certain group of six emotional facial expressions (the so-called "universal", or basic facial expressions: anger, disgust, fear, happiness, sadness, and surprise). However, other studies have shown important differences even for these six expressions across different cultural backgrounds. In addition to this debate, another important aspect that has so far been largely neglected is that human communication consists of a much broader variety of signals that do not only transport emotional, but also cognitive and other socially-regulated intentions. In order to better understand how humans perceive and process facial expressions, detailed investigations of the perceptual and cognitive aspects of facial expression processing, taking into account the broad repertoire of expression signals as well as cross-cultural contexts, are necessary.

Secondly, concerning computational aspects of facial expression research, many frameworks for automatic and efficient human-computer-interaction have been proposed by computer vision researchers. So far, however, the available frameworks are not capable of recognizing more than the basic (six) emotional facial expressions. If we want to interpret the much larger range of general conversational expressions, we need to address some crucial obstacles: first, there is a large variability across individuals in expressing certain intentions, which represents a challenge for efficient modeling of expression categories. Second, conversational signals are often conveyed using highly subtle facial movements, which is another challenge for automatic facial feature tracking algorithms. Therefore, we need to develop novel computational frameworks that are capable of dealing with these issues such that we may interpret and process the full range of human communication.

7.2 Context in Brain and Cognitive Engineering

One of the core research foci of Brain and Cognitive Engineering is to use results from cognitive neuroscience to improve human-computer-interfaces and computer algorithms in general – depending on the domain used, one may call this biologically-, perceptually-, or cognitively-motivated computer science. Conversely, novel developments in machine learning and computer science can be used to increase our understanding of fundamental perceptual and cognitive processes in the brain. The present chapter offers a perspective on this research focus: we are

first going to investigate the perceptual and cognitive aspects of facial expression processing in a cross-cultural context using state-of-the-art analysis methods. In a second step, we are then trying to design a computer vision system that takes into account some core aspects of the previously analyzed perceptual and cognitive data.

7.3 Previous Work

A series of studies conducted by Paul Ekman and colleagues in the 1960s and onwards have found evidence for the claim that certain facial expressions are "universal" across different cultures, that is, that recognizability and interpretability of these expressions is invariant across cultural contexts [1, 2]. However, recent research has cast doubts on the strong version of these statements and has shown that the concept of "universality" may be flawed, as results seem to be dependent on experimental and analysis methods [3–5]. Importantly, the aforementioned almost exclusively focused on emotional expressions only, although the full repertoire of facial expressions spans a much broader range of signals [6–9]. Indeed, relatively little is known about how we interpret and process conversational signals (such as a thoughtful or bored expression) or social signals (such as a wink, or a raised eyebrow). In addition to the cultural dependencies of emotional facial expressions, therefore the perceptual, cognitive, and cross-cultural aspects of more general facial expressions remain to be investigated [10–12].

The field of computational analysis of facial expressions has a rich history– albeit one focused again almost exclusively on the six basic facial expressions (for recent reviews, see [13, 14]). A recent recognition system achieving good recognition scores for the six emotional expressions by Kanaujia et al. [15] is based on an extended AAM (Active Appearance Model) that tracks the whole face. Many other frameworks are based on first detecting facial action units [16] (elementary muscle movements of the face) and then recognizing the six emotional expressions by detecting combinations of such facial action units [17]. These systems typically achieved the highest performance and are current state-of-the-art. Going beyond emotional expressions, Bousmalis et al. [18] tried to deal with two conversational expressions such as agreement and disagreement. In addition, McDuff et al. [19] developed an algorithm that is able to infer valence labels of continuous facial action sequences in unsegmented videos. However, most studies to date are based on rather constrained lab-settings and usually work only for a few kinds of facial expressions (mostly the six basic emotional expressions).

7.4 Database for Cross-Cultural Research

In order to conduct either cognitive neuroscience-related or computational research on conversational facial expressions, a suitable database is needed as a resource. Even when focusing on the six universal expressions, a large percentage of existing

databases consists mainly of peak frames of expressions (i.e., static images) that do not contain dynamic movements. Indeed, several studies have shown that dynamic processing of facial expressions is different both in terms of behavioral (i.e., recognition accuracy [20]) and neuroimaging components [21]. Hence, the database needs to support dynamic stimuli in order to provide ecologically valid data.

Furthermore, when thinking about the broad range of human communication, there is a lack of databases containing conversational expressions. For this reasons, a few years ago we recorded the MPI facial expression database [22] that contains video sequences of both emotional and conversational facial expressions. Recently, we complemented this resource with a Korean equivalent: the KU facial expression database [23] that was developed with the exact same protocols than the German version. The MPI facial expression database has a total of 55 different facial expressions performed by 20 native Germans, whereas the KU facial expression database contains 55 plus 7 additional (=62) facial expressions performed by 20 native Koreans. The actors were recorded with three high-resolution video cameras yielding different points of view.

In order to ensure a good compromise between fully scripted (but potentially posed and unnatural) and unscripted (natural, but non-controlled) expressions, we employed a method-acting protocol during the recordings. For this, the experimenter read a developed scenario containing a short description of an event to the actor and asked them to imagine themselves in the scenario and to react accordingly. This process was repeated three times to yield three repetitions of each expression. The scenarios were designed to accommodate a large range of different emotional and conversational contexts. Importantly, the scenarios were designed with a conceptual hierarchy in mind: for example, there are many types of smile (pure smile, sad smile, reluctant smile, flirtatious smile, . . .) or many types of agreement (pure agreement, considered agreement, reluctant agreement, . . .). Indeed, for many types of expressions we were able to find a hierarchical structure. The full list of expressions and scenarios can be found in [22].

The resulting databases comprise two large (>20,000 video sequences), fully compatible datasets recorded in different cultural contexts. Examples for three different expressions are shown in Fig. 7.1.

After developing the databases, there was a validation step with both databases using German and Korean participants: for this, video sequences from each database were given to each group of participants, and each participant was asked to name the expressions corresponding to each video sequence using less than 4 words. Three independent raters then rated the answers as valid or invalid given the scenario descriptions. Using the most conservative criterion that a sequence is only rated as valid if it is approved by all three raters, on average, the MPI database and the KU database yielded 60 % and 57 % valid sequences, respectively. Using a less strict criterion of 2 out of 3 raters, we found validity scores of 71.5 % and 66.1 %, respectively.

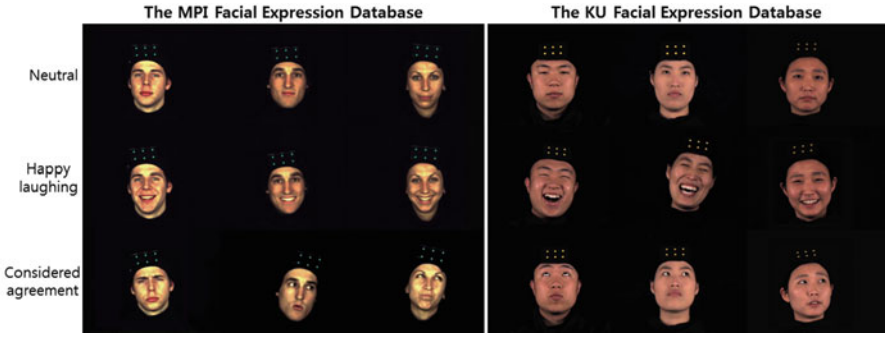

Fig. 7.1 Examples of the MPI (*left*) and KU (database) for three expressions. Note the considerable variation among individuals that is visible even in the static peak frames depicted here

7.4.1 Cognitive Study

We now turn to the first (cognitive) aspect of the present chapter in which we use the two databases to investigate the underlying dimensions of the complex space of emotional and conversational expressions in a cross-cultural context. For the experiments, 540 video sequences from the MPI facial expression database and 620 video sequences from the KU facial expression database were used as stimuli. Each group of stimuli contains expressions from 10 actors (MPI: 54 expressions of 10 actors, KU: 62 expressions of 10 actors). We conducted two fully crossed experiments across two countries, recruiting two participant groups in both Germany and Korea. For all experiments, we recruited only native German and Korean participants, where care was taken to control for exposure to the non-native cultural background (i.e., Asian/Korean for German participants, and Western/German for Korean participants). A total of 42 German participants and 44 Korean participants were recruited for this experiment.

The experiments consisted of a free-grouping task. Each group of participants received either 540 videos of the MPI facial expression database, or 620 videos of the KU facial expression database as video files in random order. Participants were then asked to group the expression sequences (i.e., to watch the video sequences and to move them into folders that they created one-by-one). There were no restrictions as to the number of clusters or the number of sequences in each cluster. In order to analyze the data, we generated confusion matrices for each of the four participant groups. Each confusion matrix tallies how often each expression was grouped with other expressions. With these matrices we then performed multidimensional scaling to identify the underlying topology and dimensionality of the resulting facial expression space.

The confusion matrices (see Fig. 7.2) showed similar structure for both databases as shown by the overall similar pattern: for example, the patterns for expressions belonging to the expression groups of “agreement” (expression labels starting with

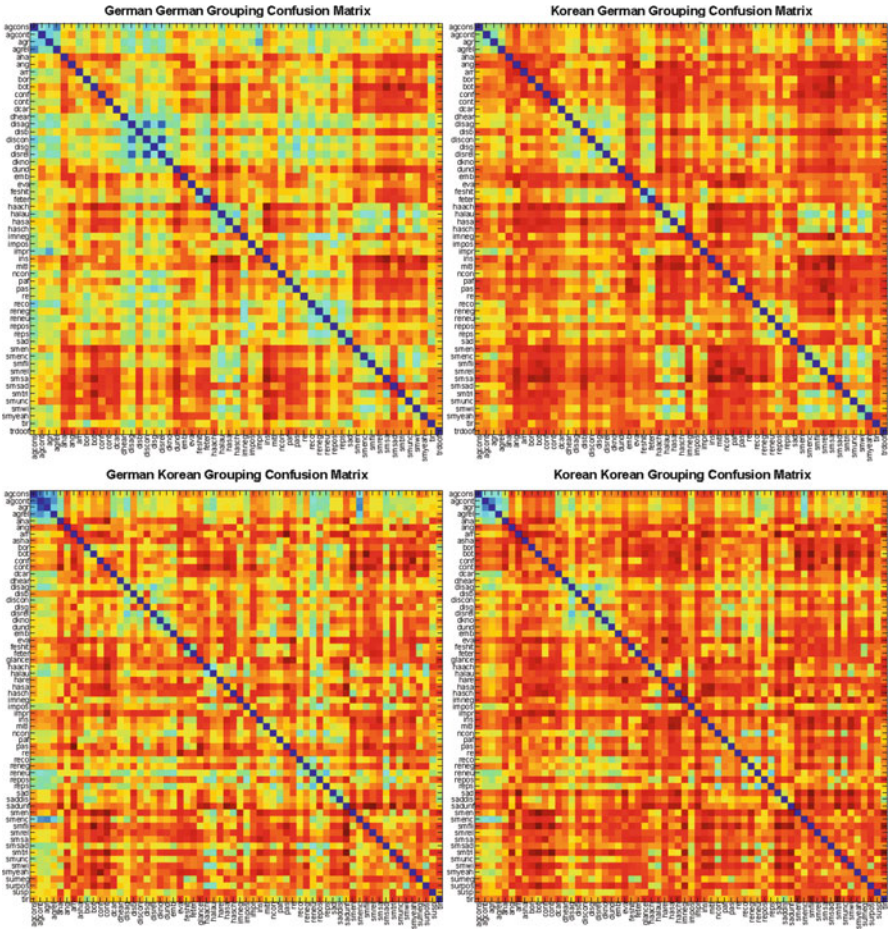

Fig. 7.2 Confusion matrices for the four participant groups. *Blue* indicates similarly grouped expressions, whereas *red* indicates dissimilarly grouped expressions

"ag" in Fig. 7.2), "disagreement" (labels starting with "disag" in Fig. 7.2), and "thinking" (labels starting with "re" in Fig. 7.2) were seen as quite similar in all confusion matrices from German and Korean participants. In contrast, the data tended to yield more confusions for non-familiar cultural judgments: expressions belonging to the basic-level groups of "smiling" and "happiness" showed more confusion for cross-cultural (i.e., Korean-German Grouping or German-Korean Grouping in Fig. 7.2) than for within-cultural judgments.

Multidimensional scaling was then used to examine the first two dimensions of the low-dimensional embedding of the grouping data (see Fig. 7.3). Comparing the positions of the expressions located at the outsides of the space (e.g., for the KU facial expression database "eva" = evasive, "impr" = impressed,

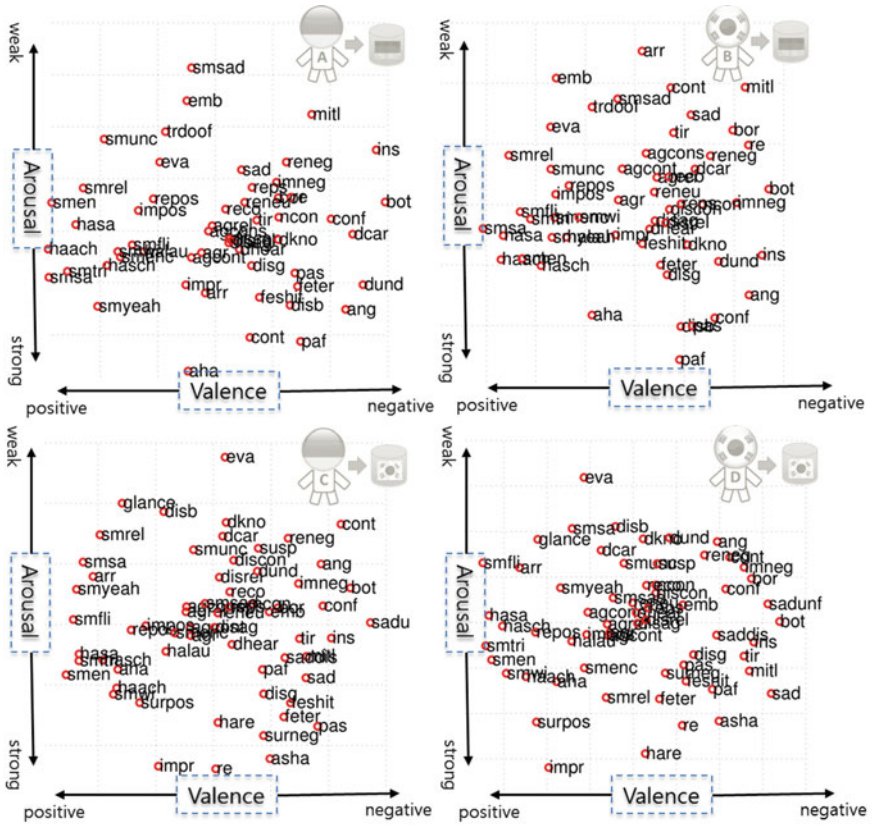

Fig. 7.3 Two-dimensional MDS solutions obtained from the four confusion matrices shown in Fig. 7.2

"smfli" = flirting smile, "bot" = bothered, and for the MPI facial expression database "emb" = embarrassed, "paf" = pain felt, "smsa" = sardonic smile, "bot" = bothered), we can see how similar the two reconstructed spaces are for each database. In addition, when comparing the KU and MPI databases, we can clearly see that expressions of the "smiling"-group (expressions starting with "sm") are located on the left side, whereas expressions such as anger ("ang") or bothered ("bot") are located on the right side of the plot. Hence this dimension recovered by the multidimensional scaling analysis corresponds to valence (positive-negative). A similar analysis reveals that the top-bottom dimension is that of arousal (weak-strong). Importantly, these dimensions are robustly recovered for both databases and both groups of participants. This shows that whereas there are differences between cultures (and to some degree, between databases), the overall structure of the space of facial expressions can be robustly explained by the two dimensions of valence and arousal.

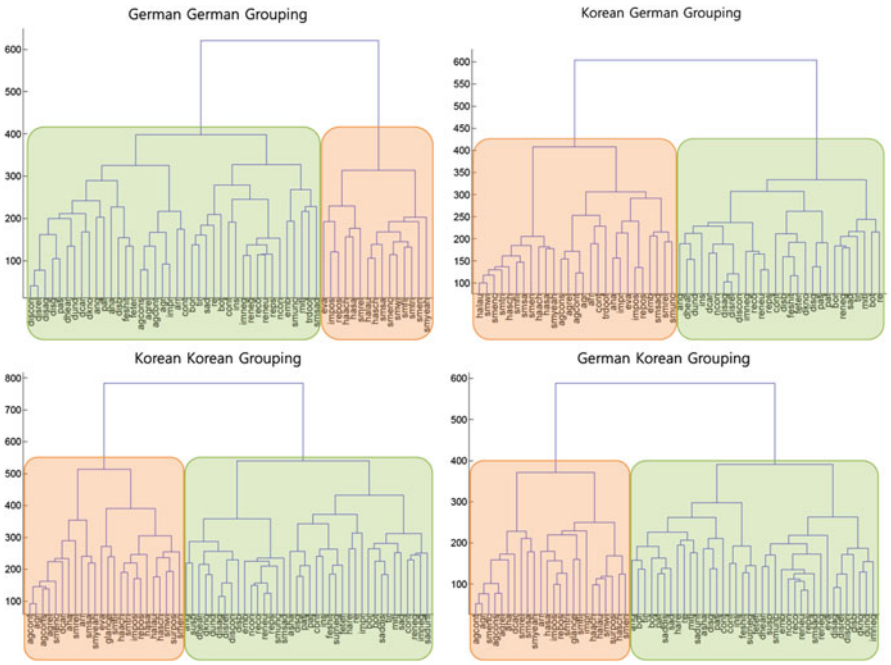

Fig. 7.4 Dendrograms from hierarchical clustering of the four participant groups. The expressions clustered in the two big groups mostly consist of positive expressions for the *red cluster* and negative expressions for the *green cluster*. The biggest difference appears for the German-German grouping, in which the "agree"-expressions (labels starting with "ag") get grouped into the *green cluster*, which does not happen for the other grouping datasets

In addition, we employed bottom-up hierarchical clustering (using the Ward-criterion) to produce a clustering view of the data. Interestingly, when looking at the resulting four dendrograms, all expressions are divided into two big clusters at the first level for the four participant groups: these clusters include for the most part valence-positive (red) and valence-negative (green) expressions (see Fig. 7.4). Furthermore, the clusters on lower levels of the hierarchy re-produce our own conceptual hierarchy well: in most cases, agree, disagree, and thinking expressions receive their own cluster, for example. Hence, this clustering indirectly validates the hierarchical structure of conversational expressions that we used also during recording and design of the database.

7.4.2 Computational Study

Closer inspection of the two databases mentioned above shows that for many types of expressions there was considerable inter-person-variability – despite clear

interpretability by human observers. Such variability will present a challenge for computational approaches. In addition, the number of categories (>50) is another issue that learning algorithms would need to deal with.

In the following, we present a computational recognition framework that tries to address these issues by using a powerful graphical sequence modeling approach: Latent Dynamic Conditional Random Fields (LDCRF). We train and test this modeling approach on the MPI database in the present chapter. For the computational experiments, we used expressions from 10 actors and took the first repetition as training data, and the second and third repetition of each expressions as testing data.

Importantly, we know that the structure of conversational facial expressions is hierarchical; for example, the expression of 'considered agreement' has two sub-expressions (considering and agreeing). In fact, these two sub-expressions can be shared across a wider range of expressions, since the "considering" part can also equally lead to a considered disagreement (another expression in the database) or simply stop without continuing (to yield thinking/considering, yet another expression in the database).

This observation also was the motivation for choosing LDCRFs for our task. Traditionally, Hidden Markov Models (HMMs) have often been used to model dynamic expressions. However, in order to predict the multiple categories of conversational expressions, Conditional Random Fields (CRFs) are more suitable because observations and latent states may follow conditional distributions. Furthermore, LDCRFs as an extension of CRFs are necessary since the hidden states (or latent factors) in LDCRFs are able to represent the required sub-expression dynamics discussed previously. Here, we compare the recognition performance of CRFs and LDCRFs in particular.

Both algorithms work on feature vectors extracted from the video sequences of the MPI facial expression database using the Computer Expression Recognition Toolbox [24]. The feature vectors include the intensity of 19 core facial action units as well as 3D-head rotations (yaw, pitch, roll). The result of tracking is used to automatically generate frame-wise labeling information. Since a manual annotation of each frame (as required by the CRF/LDCRF algorithms' training stage) would be a lot of work, we first set the intensity of the neutral expression (that by definition consisted of the first frame of each video sequence) based on the extracted feature vectors as the baseline. Each subsequent frame was then set as a non-neutral frame of the corresponding expression label, depending on a simple threshold difference to the neutral frame.

To compare the two algorithms (CRF and LDCRF), we chose to focus on six different expressions (considered agree, disagree, disgust, sad, I don't care, and happy laughing). Note that the sharing of sub-expressions here also would work in favor of the LDCRF, since sharing can of course also happen across the different actors. Accordingly, although the training time for LDCRF is longer the recognition rate is significantly higher than that of CRFs (88.6 % versus 77.1 %).

As an extension, we compared the previously mentioned human data with trained CRF models *on all expressions*. We compared confusion matrices that show how

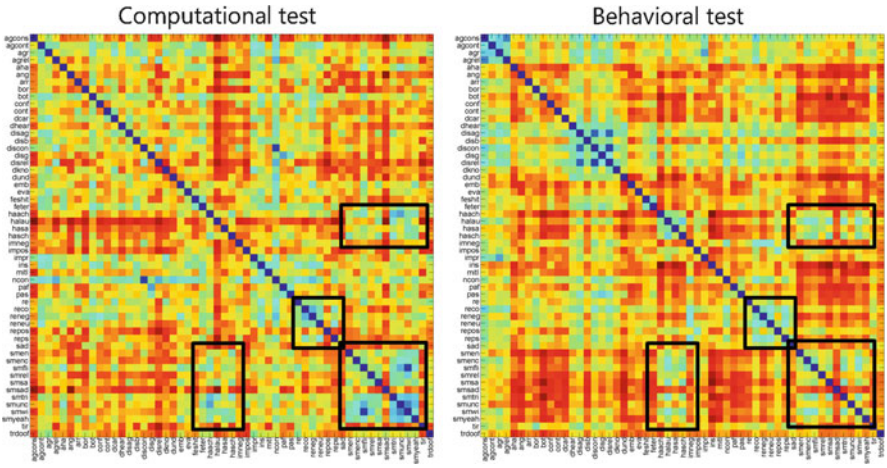

Fig. 7.5 Confusion matrices for computational (CRFs) and behavioral data (German-German grouping)

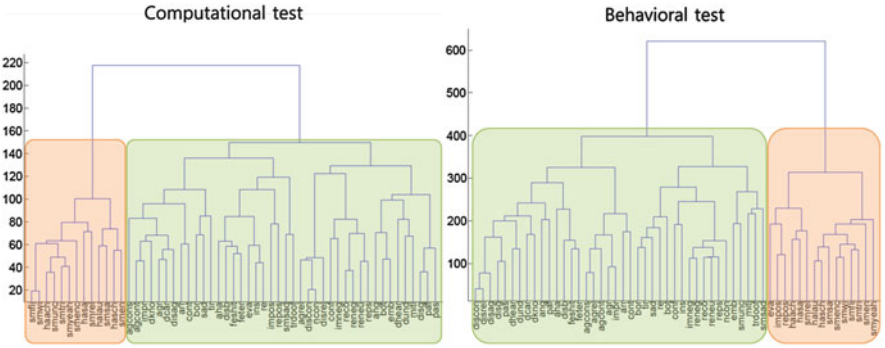

Fig. 7.6 Dendrograms for computational (CRFs) and behavioral test (German-German grouping). Note the similarity in the two larger (*red and green*) clusters between the two dendrograms – again, smiling expressions (expressions starting with "sm" or "ha") are split off first for both methods. Note, however, that clusters at lower levels of the hierarchy have differences. Cf. also Fig. 7.4 for human data

frequently expressions were confused with each other from both behavioral data and a simple computational test with the CRF models. The result is shown in Fig. 7.5. Although there are differences in the details, both matrices show a similar structure: for example, the clear block patterns in which "smiling" and "happy" categories are often confused is visible for both human and computational data (highlighted in Fig. 7.5).

This similarity is also visible in the hierarchical clustering, when comparing behavioral with computational data (see Fig. 7.6): again, the larger two clusters split off the smiling (valence-positive) expressions first. At the lower levels, however,

differences start to appear: agree, and disagree expressions, for example, do not get grouped together in the computational clustering, whereas they clearly do in the behavioral data (see Fig. 7.4).

Overall, these results imply that relatively simple graphical models of computationally extracted features are able to replicate some broad-scale patterns of human performance.

7.5 Discussion

Using two large databases from two different cultural contexts, we first investigated cross-cultural perception of facial expressions – using expressions containing not only emotional, but also conversational aspects of facial expressions shown as dynamic data. We conclude that although expressions from a familiar background are more effectively grouped (i.e., less confused), the evaluative dimensions for both German and Korean cultural contexts are exactly the same, showing that cultural universals exist even in this complex space. The next step will consist of running rating experiments on a variety of conceptual scales to correlate the *implicit* dimensions obtained here (valence and arousal) to that of *explicit* judgments. Additional research will be conducted to extend this experiment to a larger participant base using crowd-sourcing to investigate different cultural backgrounds as well as dimensions of age.

For the computational study, we showed that since conversational expressions contain a hierarchical structure, modeling that takes into account this structure (LDCRF) shows a considerable advantage in recognition rates even on the smaller number of expressions tested here. In addition, we showed that the graphical models such as conditional random fields yield confusion patterns similar to those of human grouping on a broad scale. Future research will need to use more data from the full database (20 actors) to develop better models of facial expressions. Since such training is very costly at present with the extended CRF models, more efficient training algorithms will need to be developed as well to cope with the large amounts of data.

References

1. Ekman P (1994) Strong evidence for universals in facial expressions. Psychol Bull 115(2):268–287
2. Izard CE (1994) Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychol Bull 115(2):288–299
3. Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol Bull 115(1):102–141
4. Nelson NL, Russell JA (2013) Universality revisited. Emot Rev 5(1):8–15

5. Jack RE, Garrod OGB, Schyns PG (2013) Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Curr Biol 24(2):187–192
6. Lee K-U, Khang HS, Kim K-T, Kim Y-J, Kweon Y-S, Shin Y-W, Liberzon I (2008) Distinct processing of facial emotion of own-race versus other-race. NeuroReport 19(10):1021–1025
7. Matsumoto D, Nakagawa S, Estrada A (2009) The role of dispositional traits in accounting for country and ethnic group differences on adjustment. J Pers 77(1):177–211
8. Jack RE, Blais C, Scheepers C, Schyns PG, Caldara R (2009) Cultural confusions show that facial expressions are not universal. Curr Biol 19(18):1543–1548
9. Jack RE, Garrod OGB, Yu H, Caldara R, Schyns PG (2012) Facial expressions of emotion are not culturally universal. Proc Natl Acad Sci 109(19):4–7
10. Schmidt KL, Cohn JF (2001) Human facial expressions as adaptations: evolutionary questions in facial expression research. Am J Phys Anthropol 33(S33):3–24
11. Elfenbein HA, Beaupré M, Lévesque M, Hess U (2007) Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. Emotion 7(1):131–146
12. McCarthy A, Lee K, Itakura S, Muir DW (2008) Gaze display when thinking depends on culture and context. J Cross Cult Psychol 39(6):716–729
13. Gatica-Perez D (2009) Automatic nonverbal analysis of social interaction in small groups: a review. Image Vis Comput 27(12):1775–1787
14. Vinciarelli A, Pantic M, Heylen D, Pelachaud C, Poggi I, D'Errico F, Schröder M (2012) Bridging the gap between social animal and unsocial machine: a survey of social signal processing. IEEE Trans Affect Comput 3(1):69–87
15. Kanaujia A, Metaxas D (2006) Recognizing facial expressions by tracking feature shapes. In: Proceedings – International conference on pattern recognition. Hongkong, vol 2, pp 33–38
16. Rivera J, Kreuz T (2009) Reading faces with conditional random fields, Technical report. Robotics Institute, Carnegie Mellon University, Pittsburgh
17. Chang KY, Liu TL, Lai SH (2009) Learning partially-observed hidden conditional random fields for facial expression recognition. In: 2009 IEEE computer society conference on computer vision and pattern recognition workshops. Miami, pp 533–540
18. Bousmalis K, Morency LP, Pantic M (2011) Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition. In: 2011 IEEE International conference on automatic face and gesture recognition and workshops. Santa Barbara, pp 746–752
19. McDuff D, El Kaliouby R, Kassam K, Picard R (2010) Affect valence inference from facial action unit spectrograms. In: 2010 IEEE computer society conference on computer vision and pattern recognition – workshops. San Francisco, pp 17–24
20. Cunningham DW, Wallraven C (2009) Dynamic information for the recognition of conversational expressions. J Vis 9(13):7.1–17
21. Bülthoff HH, Cunningham DW, Wallraven C (2011) Dynamic aspects of face processing in humans. In: Li SZ, Jain KA (eds) Handbook of face recognition. Springer, London, pp 571–576
22. Kaulard K, Cunningham DW, Bülthoff HH, Wallraven C (2012) The MPI facial expression database – a validated database of emotional and conversational facial expressions. PLoS One 7(3):e32321
23. Lee H, Shin A, Kim B, Wallraven C (2012) The KU facial expression database: a validated database of emotional and conversational expressions. In: Proceedings of Asian Pacific conference on vision. Incheon
24. Bartlett M, Littlewort G, Wu T, Movellan J (2008) Computer expression recognition toolbox (CERT). In: 2008 8th IEEE International conference on automatic face and gesture recognition. Amsterdam

Chapter 8
Bottom-Up Processing in Complex Scenes: A Unifying Perspective on Segmentation, Fixation Saliency, Candidate Regions, Base-Detail Decomposition, and Image Enhancement

Boyan Bonev and Alan L. Yuille

Abstract Early visual processing should offer efficient bottom-up mechanisms aiming to simplify visual information, enhance it, and direct attention to make high-level processing more efficient. Based on these considerations, we propose a unified approach which addresses a set of fundamental early visual processes: segmentation, candidate regions, base-detail decomposition, image enhancement, and saliency for fixations prediction. We argue that for complex scenes all these processes require hierarchical segmentwise processing. Furthermore, we argue that some of these visual tasks require the ability to decompose the appearance of the segments into "base" appearance and "detail" appearance. An important, and surprising, result of this decomposition is a novel method for successfully predicting human eye fixations. Our hypothesis is that we fixate on segments that are not easy to model, e.g., are small but have a lot of detail, in order to obtain a higher resolution representation for further analysis. We show performances on psychophysics data on the Pascal VOC dataset, whose images are non-iconic and particularly difficult for the state-of-the-art saliency algorithms.

Keywords Bottom-up visual processing • Image segmentation • Base-detail decomposition • Saliency

8.1 Introduction

Low-level vision is visual processing that treats images as patterns and makes no specific assumptions about the objects that might be present or the structure of the scene. In short, the processing is generic and intended to be suitable for all images, regardless of their semantic content or high level layout. Examples of

B. Bonev (✉) • A.L. Yuille
Department of Statistics, University of California, Los Angeles, CA, USA
e-mail: bonev@ucla.edu; yuille@stat.ucla.edu

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_8

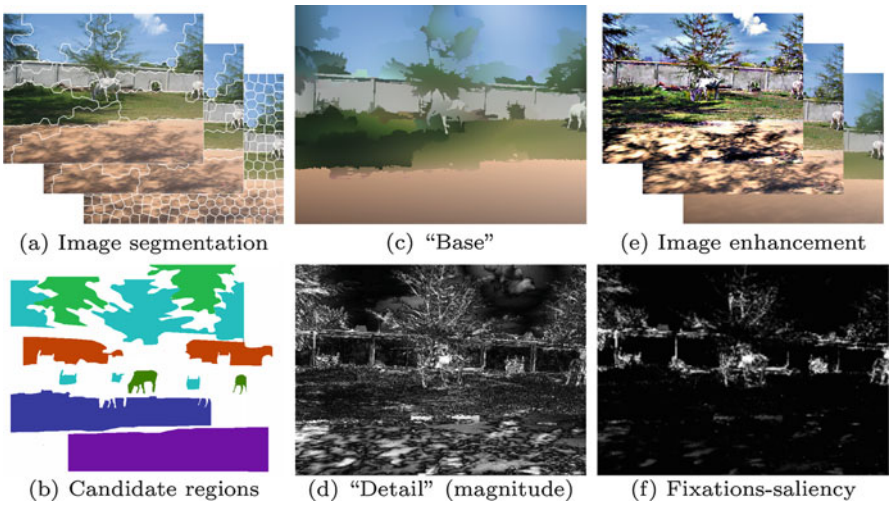

Fig. 8.1 We propose a unified approach for several low-level visual processes: (**a**) image segmentation – a hierarchy of image partitions at multiple levels; (**b**) candidate regions – a pool of possibly overlapping proposals for further study by object recognition methods (best candidates illustrated); (**c**,**d**) "base-detail" decomposition – expressing the image as the sum of a non-local smooth appearance term and s residual, or detail, which captures the texture patterns; (**e**) image enhancement – controlling the amount of detail in the image; (**f**) saliency for fixations prediction – a model predicting bottom-up human visual attention

low-level vision tasks include segmentation, candidate regions or object proposals, and image enhancement. Low-level processing is typically performed in preparation for high-level tasks, and is used to allocate computational resources for more detailed processing. In mammalian visual systems low-level vision is believed to be performed in the retina and area V1 of the visual cortex.

In this paper we propose a unified framework for several low-level vision tasks (Fig. 8.1) that are typically modeled separately. These tasks include the generation of hierarchical image segmentations, proposing candidate regions for object detection and recognition, base-detail decomposition – where an image is decomposed into a visual summary plus fine details – image enhancement and the prediction of human eye fixations.

We start by producing a hierarchical decomposition of the image into segments which have roughly uniform homogeneity as measured by texture and color cues. Segments at higher levels of the hierarchy are generally larger and less homogeneous. But in our approach, it is important that the size of segments within each level of the hierarchy have different sizes because some image regions (e.g., sky) are much more homogeneous than others (e.g., a road containing several cars).

Different segments of the hierarchy are combined into groups of up to three to make proposals for the positions and shapes of objects and background "stuff" [10], which we refer to as candidate regions. They consist of a pool of 500–1500 regions which are later evaluated by a high-level method, which is out of the scope of this

paper. The high-level method computes category-specific scores to identify regions which correspond to object or background categories.

We define "base-detail" decomposition as the separation of the image into a coarse description of the image appearance, and a description containing the texture and details. The image is the sum of both. More precisely, the base is obtained by fitting smooth appearance models (polynomials) to the image segments and the detail is the residual. For examples, see Figs. 8.1c, d and 8.8. This base-detail decomposition enables us to process the image in several ways, such as enhancing the details and/or the base. For example, we can remove the shadows (details) from a grass lawn (base). Surprisingly, as we now discuss, we can use base-detail decomposition to predict human eye fixations for free viewing.

It is well known that when humans examine an image they do not gaze on it uniformly but instead they fixate on certain parts of the image. The fixation saliency model we propose favors small segments which have strong details. This has the following intuition: large segments are typically homogeneous regions (e.g., sky, water, or grass) which may be easily processed (i.e., classifying these regions may be easy using methods which use summary image statistics and do not model the detailed spatial relations). The detail is less important in the large segments but in small segments the detail may correspond to structures which require more detailed models to process. We describe experiments showing that our fixation saliency model predicts human fixations with a state-of-the-art performance on complex datasets, like Pascal [17] and Judd [31].

Our work is motivated both by attempts to understand how primate visual systems work and by efforts to design computer vision systems with similar abilities. We provide a computational model for performing these visual tasks but in this paper we do not develop any detailed biological evidence for this theory. Instead we concentrate on performance on complex visual scenes, instead of artificial stimuli, because we think it is important to model visual abilities in real-world conditions.

8.2 Background and Related Literature

There is an enormous literature on segmentation much of it using Markov Random Field (MRF) models [22]. Our work follows the alternative strategy of decomposing images into subregions which have roughly similar statistical image properties [1, 33, 45, 52]. There are a variety of hierarchical approaches which exploit the intuition that image structures occur at different scales and that multi-scale is required to capture long-range interactions within the images [19, 53]. Our approach to hierarchies follows the strategy of starting with an over-segmentation of the image, produced by an efficient algorithm like [1], followed by recursive grouping to get larger segments at different levels of the hierarchy [4]. This relates closely to Segmentation by Weighted Aggregation [20], a recent variant [3], and extensions to video segmentation [48].

Detecting candidate regions, which make proposals for the positions and sizes of objects, is a new but increasingly important topic in computer vision. This is because it offers an efficient way to apply powerful methods, such as Deep Convolutional Neural Networks (DCNN) [34], to detect and recognize objects in images. Instead of needing to apply DCNNs exhaustively, at every image position and scale, it is only necessary to apply them to a limited number of candidate regions. Our method for detecting candidate regions differs from existing methods because we propose regions for both objects and background regions or "stuff" (e.g., sky). Recent work on detecting candidate regions includes methods which group segments into combinations [5, 6]. Most methods in the literature have been evaluated for finding segments which cover foreground objects [46], while ours detects background classes as well. Finally, there are other methods which differ in that they mainly exploit the edges instead of the appearance statistics [15, 29, 55]. We should also mention hierarchical segmentation which has been used to learn models of objects [43].

There is no existing work that directly addresses "base-detail" segmentation, but there is a large literature on closely related topics. In the digital image processing community there is a related concept, "base-detail separation", but it is performed locally [7] by applying bilateral filters. A related topic is gain control which has been studied in primate visual systems, particularly in the retina, and seeks to compress the dynamic range of the input intensity while preserving the local contrast and detail [14, 42]. We note that detection of detail is also at the heart of many super-resolution methods [54] and it is related to image enhancement. Enhancement approaches do not typically use segment-based methods [25, 41] and instead use local methods like the bilateral filter in [7] or the weighted least squares [18]. There are some exceptions, like [49] where segment-wise exposure correction is proposed.

Another related topic is work in the shape from shading community, where intensity patterns are decomposed into smoothly varying shading patterns and more variable texture/albedo components [9, 26, 27] (here the base roughly corresponds to the shading and the texture/albedo to the detail). Researchers in shape from shading make prior assumptions for performing the decomposition into shading and texture/albedo [8] (which are not needed if the same object is viewed under different lighting conditions [47]). Similar decompositions assumptions are also applied to the classic Mondrian problem [32].

Predicting human eye fixations is a long studied research topic [30]. In this paper we address only bottom-up saliency prediction, as performed in a free-viewing task, and do not consider top-down processes involving which involve cognitive factors, e.g., eye fixations when performing a task such as counting the animals in an image. One of the first successful methods for predicting human eye fixations was Itti's original model [30]. Image signature is a simple method which give good results [28] and other recent methods are reviewed in [11]. The most successful current method is Adaptive Whitening Saliency [21] and we make comparisons to it in our experiments. Finally, there are other works [37, 38] which studies the saliency of visual objects and use candidate regions to make predictions [13]. Objects are judged to be salient based on the number of eye fixations which occur within them.

By contrast, fixation saliency only predicts positions and outputs a fixations map (in Sect. 8.4.3, see Fig. 8.16, second column). The eye fixation saliency models we propose is based on base-detail decomposition, which makes it substantially different from any method in the literature. Our experiments show it performs at the state of the art.

Finally, although biology is out of the scope of this paper, we find it interesting that recent biological vision studies suggest that early visual processing is more sophisticated than traditional models of the retina and V1, which mainly emphasize linear spatiotemporal filters. For example, studies of the retina suggest that it is "smarter than scientists believed" [23] and contains a range of non-linear mechanisms which might perhaps be able to implement parts of the theory of theory we propose here. Moreover, there is growing appreciation of the richness of computations that can be performed in area V1 of the visual cortex, including possibly fixation saliency [51].

8.3 Method

In this section, we describe the details of the proposed approach. We address a set of fundamental low-level vision processes: segmentation, candidate regions and salient objects proposals, base-detail decomposition, image enhancement, and bottom-up saliency. Instead of being treated as separate tasks, we address them in terms of a unified approach of bottom-up vision processing.

8.3.1 Segmentation: Hierarchical Image Partitioning

Image segmentation is a classic task of low-level vision. But in this paper we do not consider segmentation as a goal in itself. Instead, we seek to obtain a hierarchy of segmentations, or partitions of the image into segments, which can be used as components for other processing, as will be described in the next subsections.

An image partition is a decomposition of the image into non-overlapping subregions, or *segments*. More formally, we decompose the image lattice $\mathcal{D}$ into a set of segments $\{\mathcal{D}_i : i = 1, \ldots, n\}$ such that:

$$\mathcal{D} = \bigcup_{i=1}^{n} \mathcal{D}_i, \quad \text{s.t. } \mathcal{D}_i \bigcap \mathcal{D}_j = \emptyset, \ \forall i \neq j.$$

A hierarchical partition of an image is a set of decompositions indexed by hierarchy level $h = 1, \ldots, H$. Each level gives an image partition $\mathcal{D} = \bigcup_{i=1}^{n_h} \mathcal{D}_i^h$, where n_h is the number of segments in the partition at level h. The decompositions are *nested* so that a segment $\mathcal{D}_i^h$ at the hierarchy level h is the union of a subset of segments

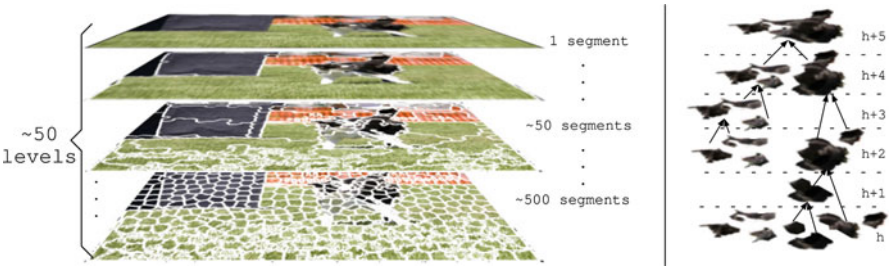

Fig. 8.2 *Left*: Multiple levels in a hierarchy. Segments with a good coverage of objects or parts may happen at different levels. 80 % to 90 % of the segments can be discarded because they go across boundaries of objects or because they don't cover a large area of an object. *Right*: Segments at level $h+1$ are composed of one or two segments in level h

at the previous level $h-1$, so that $\mathcal{D}_i^h = \bigcup_{j \in Ch(\mathcal{D}_i^h)} \mathcal{D}_j^{h-1}$, where $Ch(\mathcal{D}_i^h)$ denotes the child segments of segment i at level h (in this paper each segment is constrained to have at most two immediate children, see Fig. 8.2-right). This enables us, by recursion, to express a segment in terms of compositions of its descendants in many different ways. In particular, we can decompose a segment into its descendants at the first level, $\mathcal{D}_i^h = \bigcup_{j \in Des(\mathcal{D}_i^h)} \mathcal{D}_j^1$. This hierarchical structure is common in the segmentation literature, for example in [4]. Figure 8.2 illustrates the hierarchical partitioning of an image.

In this paper, our hierarchical partitioning is designed based on the following related considerations. Firstly, we prefer segments to have roughly homogeneous image properties, or *statistics* $\vec{S}$ (e.g., color/texture/detail) at each level, which means that segments at the same level can vary greatly in size (e.g., segments on the grass in Fig. 8.2 will tend to be larger than in less homogeneous regions of the image, like the dog). Secondly, segments at higher levels should be less homogeneous because they are capturing larger image structures (e.g., by merging more homogeneous image structures together). Thirdly, segments are likely to have edges (i.e., image intensity discontinuities) near their boundaries. Fourthly, we want an efficient algorithm which can dynamically compute this hierarchy using local operations by merging/grouping segments at level $h-1$ to compose larger segments at level h.

Our work is guided by standard criteria for image segmentation [33, 45, 52] which propose minimizing a cost function of form:

$$E(\{\mathcal{D}_i\}, \{\vec{S}_i\}) = \sum_i \sum_{x \in \mathcal{D}_i} |\vec{S}_i - \vec{S}(x)|^2 - \lambda \sum_i \sum_{x \in \partial D_i} e(x). \tag{8.1}$$

Here $\vec{S}(x)$ denotes image statistics at position x (e.g., color, texture features), $\vec{S}_i$ is summary statistics of the region i, λ is a non-negative constant, and $e(x)$ is a measure of edge strength (taking large values at image discontinuities), and ∂D_i is the boundary of segment $\mathcal{D}_i$.

We initialize our algorithm by using the SLIC [1] algorithm to compute the lowest level, $h = 1$, of our hierarchy. Essentially, SLIC performs an expectation-minimization of (8.1) for a fixed number n_1 of segments. It uses the color and position as statistics, without including an edge term, that is, $\lambda = 0$ in (8.1). More precisely, $\vec{S}(x) = (l(x), a(x), b(x), x)$, where l, a, b specify color channels of the Lab color opponent space and x denotes 2D spatial position.

Next, we proceed to construct the hierarchy by grouping/merging segments which have similar image statistics. The statistics are extended to include texture, shape of segments, and the variance of color (we do not use these statistics at the bottom-level because the segments are too small to compute them reliably). More precisely, $\vec{S}$ is given by the mean and the standard deviation of the Lab color space components and the first and second derivatives of the l channel, $(l, a, b, \nabla_x l, \nabla_y l, \nabla_x^2 l, \nabla_y^2 l)$, the centroids of the segment and dimensions of its bounding box (c_x, c_y, d_w, d_h). When performing merging, we use an asymmetric criterion which requires comparing the difference between the statistics of the union of the two segments i and j, $\vec{S}_{i \bigcup j}$, and the statistics of its segments $\vec{S}_i, \vec{S}_j$, that is, $||\vec{S}_{i \bigcup j} - \vec{S}_i||$ and $||\vec{S}_{i, \bigcup j} - \vec{S}_j||$. This is because our segments are allowed to have different sizes and we want to discourage bigger segments from merging with smaller segments if this will change much the statistics of one of them. Intuitively, a big segment is likely to have little change on its statistics by merging to a small one, but we want to ensure that the small one does not undergo a big change in its statistics. At each level of the hierarchy we allow the top-ranked 30 % segments to merge to another segment (rank is based on asymmetric criterion described above and prioritizes similar segments) but prevent merges where the asymmetric condition is violated. Merging is allowed between 1st and 2nd neighbors only. The precise details are described in [10].

The output is a hierarchical partition of the image. It is expressed as a set of segments $\{\mathcal{D}_i^h\}$, $1 \leq h \leq H$, $1 \leq i \leq n_h$, where h is the hierarchy level. At the highest level, $n_H = 1$. Each image region $\mathcal{D}_i^h$ and has statistics $\vec{S}_i^h$. Each level h gives a partition of the image $\mathcal{D} = \bigcup_{i=1}^{n_h} \mathcal{D}_i^h$. Each segment is composed of a set of child segments, $\mathcal{D}_i^h = \bigcup_{j \in Ch(\mathcal{D}_i^h)} \mathcal{D}_j^{h-1}$. Each segment can also be associated to its descendant segments at the $h = 1$ level: $\mathcal{D}_i^h = \bigcup_{j \in Des(\mathcal{D}_i^h)} \mathcal{D}_j^1$. This hierarchical partition can be used directly for image segmentation but, in the spirit of this paper, we think of it as a representation that can be used to address several different visual tasks as we will describe in the next few sections.

8.3.2 Candidate Regions

This section shows how to use the hierarchical partition to obtain candidate regions, or proposals, for both foreground objects and background regions, or "stuff" (e.g., sky, water, grass). Proposing candidate regions enables algorithms to concentrate computational resources, e.g., deep networks, at a limited number of locations (and

sizes) in images (instead of having to search for objects at all positions and at all scales). It also relates to the study of *salient objects* [2, 37], where psychophysical studies show that humans have tendencies to look at salient objects [16]. Note that salient objects, however, do not predict human eye fixations well [12] and these can be better described by bottom-up saliency cues [30] in a free-viewing task. However, methods that combine bottom-up saliency cues with proposals for candidate regions do perform well for both predicting human eye fixations and for the detection of salient objects [38].

We create candidate region proposals by the following strategy. Firstly, we select a subset of *selected segments* from the hierarchical partition of the image. These segments are chosen to be roughly homogeneous but as large as possible. Secondly, we make compositions of up to three selected segments to form a candidate region. These compositions obey simple geometric constraints (proximity and similarity of size). The intuition for our approach is that many foreground objects and background "stuff", can be roughly modeled by three segments or less, see Fig. 8.4. This intuition was validated [10] using the extended labeling of Pascal VOC [40] which contained per-pixel labels of 57 objects and "stuff".

The selected segments are chosen by computing the entropy gain of the combination of two child segments into their parent segment. If the entropy gain is small, then we do not select the child segments because this is evidence that they are part of a larger entity. But if the entropy gain is large, then we add the child segments to our set of *selected segments*. More precisely, we establish a constant threshold G for the entropy gain g after merging two segments $\mathcal{D}_i^h, \mathcal{D}_j^h$ into their parent $\mathcal{D}_m^{h+1} = \mathcal{D}_i^h \bigcup \mathcal{D}_j^h$. The entropy gain is defined to be:

$$g = \mathcal{H}(\mathcal{D}_m^{h+1}) - \left\{\mathcal{H}(\mathcal{D}_i^h) + \mathcal{H}(\mathcal{D}_j^h)\right\}. \tag{8.2}$$

Here $\mathcal{H}(\mathcal{D}_i^h)$ is the entropy of a segment i at level h, computed from the statistics $\{\vec{S}_k^1\}$, $k \in Des(\mathcal{D}_m^h)$ of its descendant segments at level $h = 1$ (Fig. 8.3). The entropy is computed in a non-parametric manner [10] using the approximation proposed in [35]. See an example of triplets of selected-segments in Fig. 8.4.

Fig. 8.3 *Entropy gain* (Sect. 8.3.2): When segments **a** and **b** are merged, the increase of entropy is not as big as if they were merged with **c**. *Homogeneity criterion* (Sect. 8.3.3.1): Segment **c** is homogeneous. It presents smooth variation due to shading and lighting. Segments **a** and **b** are not homogeneous. Both entropy and homogeneity are calculated from the small (first level $\mathcal{D}_i^1$) segments, illustrated with white contours

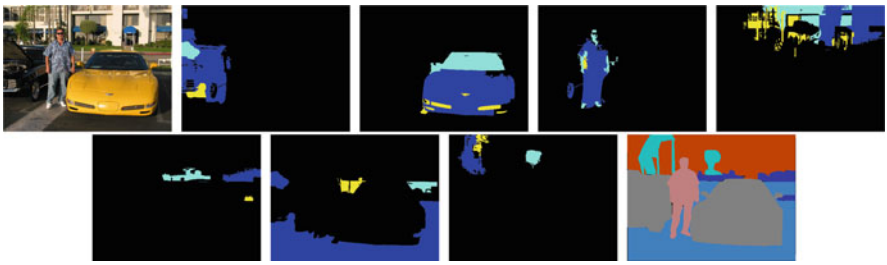

Fig. 8.4 Examples of candidate regions for foreground and background regions. Left-to-right and top-to-bottom: image, top three selected-segments for left car, right car, person, building, grass, ground, trees, and ground truth. Most objects are covered well by two to three selected-segments

8.3.3 Base-Detail Decomposition

This section analyzes the image intensities within the segments by decomposing the image into base and detail. The base $B(x)$ component is the approximate color of the region, and is required to be spatially smooth. The detail $R(x)$ is the residual $R(x) = I(x) - B(x)$ and can contain general texture, such as the patterns of grass on a lawn, or structured detail such as the writing on the label of a wine bottle.

Base-detail relates to several well studied phenomena. Firstly, it is similar to the task of preserving image contrast (i.e. the detail) performed by the early visual system when doing gain control. Secondly, it relates to the decomposition $I(x) = a(x) \cdot \vec{n}(x) \cdot \vec{s}(x)$ of images into albedo, normals and illumination when computing intrinsic images or the 2.5D sketch. But this decomposition is higher-level, relying on concepts like geometry and lighting sources, while we are modeling at a lower level. We note that in some special situations the base and the detail of a segment may correspond to the shading and the albedo of an object. Thirdly, base-detail also relates to transparency – e.g., the viewing of images through a dirty window (the dirt is the detail) – or when there is partial occlusion like tree leaves in front of a building (leaves are details). More generally, within image regions there is base appearance which changes smoothly within segments and detail which changes in a more jagged manner. This differs from the base-detail separation [7] studies in the image processing literature, which is obtained by local smoothing methods and not in a segment-wise manner.

We address base-detail decomposition in two steps. Firstly, we seek a segmentation of the image into regions which are as homogeneous and as large as possible. This is done by selecting a subset of those hierarchy segments $\{\mathcal{D}_i^h\}$ which are *maximally large and homogeneous* and form a partition of the image. Note that this includes segments at different levels h of the hierarchy. Secondly, within each segment we fit a low-order polynomial to the color intensities and define the best fit polynomial to be the base (see Sect. 8.3.3.2). We obtain the detail by computing the residual between the image and the base.

8.3.3.1 Finding Maximally Large Homogeneous Segments

Here we present a criterion for selecting non-overlapping segments from the hierarchy (while in Sect. 8.3.2 we presented a way to select overlapping segments from the hierarchy). We start from the segmentation hierarchy $\{\mathcal{D}_i^h\}$ defined in Sect. 8.3.1. We define the heterogeneity of a segment $\mathcal{D}_i^h$ by the maximum difference of the statistics of its neighboring descendant nodes at level $h = 1$. More precisely, we define the heterogeneity of segment $\mathcal{D}_i^h$ to be:

$$\max_{j,k \in Des(\mathcal{D}_i^h)} ||\vec{S}_j^1 - \vec{S}_k^1||, \ \forall\ d_G(j,k) \leq 2, \tag{8.3}$$

where $d_G(j,k)$ is the graph distance between j,k at level $h = 1$ (i.e., we evaluate only the 1st and 2nd neighbors). This criterion considers homogeneous those segments whose statistics at level $h = 1$ change smoothly across the segment. This typically happens in large segments like sky, roads, animals. Heterogeneous segments will be those which have an abrupt change in their statistics.

We then fix a threshold t_{max} and generate an image partition

$$p_{t_{max}}(I(x)) \subset \{\mathcal{D}_i^h\}, \tag{8.4}$$

containing the biggest segments whose heterogeneity is less than t_{max}. This can be done by starting at the top-level $h = H$, keeping any node whose heterogeneity is less than t_{max}, proceeding to the child nodes otherwise, and continuing down the hierarchy until we reach levels where the heterogeneity threshold is achieved. Thus, the result is a set of non-overlapping segments covering the whole image space. Note that this is different from the entropy gain criterion used in Sect. 8.3.2, which allows to select overlapping segments, as interesting structures can happen at different levels (e.g., windows as a subpart of house).

8.3.3.2 Base Modeling and Detail

We assume that the image can be expressed as $I(x) = B(x) + R(x)$ where x is 2D position, $B(x)$ is base and $R(x)$ is detail (residual of the base). Both of them include all image channels. We assume that the base is spatially smooth within each maximally large homogeneous segment and, in particular, that its color intensity can be modeled by a low-order polynomial. We make no assumption about the spatial form of the detail. (Note that for intrinsic images it is typically assumed that the shadows are spatially smooth while the texture/albedo is more jagged.)

More precisely, we define the base color of a segment by a polynomial approximation $b_k(\vec{x}_i, \vec{\omega})$ of order k, where $k \leq 3$. See examples in Fig. 8.5. We apply the polynomial approximation on each channel separately. The number of parameters $\vec{\omega}$ depends on the order of the polynomial and we use model selection to decide the

Fig. 8.5 Examples of polynomial base approximations. *Left*: original. *Center*: 0-order approximation (i.e., mean). *Right*: 0-order to 3rd-order approximation

order for each segment (we must avoid fitting a high-order polynomial to a small segment). These polynomial approximations are of form:

$$
\begin{aligned}
& b_k(\vec{x},\vec{\omega}) = \vec{x}^T\vec{\omega}, \qquad (8.5)\\
& k=0:\ \vec{x}=1,\ \vec{\omega}=\omega_0\\
& k=1:\ \vec{x}=[1,x_1,x_2],\ \vec{\omega}=[\omega_0,\omega_1,\omega_2]\\
& k=2:\ \vec{x}=[1,x_1,x_2,x_1^2,x_2^2,x_1x_2],\ \vec{\omega}=[\omega_0,\cdots,\omega_5]\\
& k=3:\ \vec{x}=[1,x_1,x_2,x_1^2,x_2^2,x_1x_2,x_1^3x_2^3,x_1x_2^2,x_2x_1^2],\ \vec{\omega}=[\omega_0,\cdots,\omega_9]
\end{aligned}
$$

The estimation of the parameters $\vec{\omega}$ of the polynomial is performed by linear least squares QR factorization [24]. The order k is selected based on the error, with a regularization term biasing towards lower order. See Fig. 8.6-right. The regularization is weighted by ζ, whose value is not critical (it is set to produce models of all orders k, and not only $k=3$). In a given segment we have a set of pixels with 2D positions x and color intensity values $I_c(x)$. For a given channel c of the segment $\mathcal{D}_i^h$, we minimize:

$$
\min_{\vec{\omega},k} \sum_{x\in\mathcal{D}_i^h} \left(I_c(x) - b_k(\vec{x},\vec{\omega})\right)^2 + \zeta k. \qquad (8.6)
$$

Fig. 8.6 Different segments can have different polynomial order k. *Left*: original. *Center*: polynomial base approximation. *Right*: order of the polynomial, where: *dark-blue*: $k = 0$, *light-blue*: $k = 1$, *yellow*: $k = 2$; *red*: $k = 3$

Fig. 8.7 Example of detail (*right image*) in front of different appearance segments: sky, road, and building

We estimate the base $B_c(x)$ of each color channel c for the whole image by fitting the polynomial for each maximally large homogeneous segment. Then, we estimate the detail to be the residual $R_c(x) = I_c(x) - B_c(x)$.

Our current method works well in most cases, see Fig. 8.7, but it is not appropriate for segments where the amount of detail is similar to the amount of base appearance. This happens, for example, for an image of a leafy tree with blue sky behind it. Such situations require a more complex model which has a prior on the details and allows the base to be fit by a more flexible function (but still smooth).

"Base-detail" provides a unified model for several visual tasks that are often modeled separately. These include: (I) Elementary tasks such as gain control, which converts the large dynamic range of luminances into a smaller range of intensities which can be encoded by neurons and transmitted to the visual cortex. A standard hypothesis is that it is performed by ganglion cells in the retina, by Difference of Gaussian, or Laplacian of Gaussian [39], filters to preserve the contrast while removing the base. From our perspective, the contrast is the detail. (II) Decomposition of intensity into albedo and shading patterns as required by shape from shading algorithms [26, 27] when used to construct the 2 1/2 sketch [39] or intrinsic image [9]. The difference is that we do not estimate 3D geometry, noting that intrinsic image models make strong assumptions about images which are often

invalid (e.g., smooth intensity patterns can be due to light sources at finite distance and not to the geometry of the viewed surface). (III) Separation of texture from background. Here the detail represents the texture patterns, e.g., the blades of grass while the base is a smooth green intensity pattern. (IV) Decomposing images into frequency components. In this case, the detail is analogous to the high-frequencies. But frequency analysis is based on linear analysis of images while our approach is inherently nonlinear because it involves segmentation. (V) Image compression. Base details suggests a strategy where the base efficiently encodes the rough appearance and the detail encodes the rest. It captures the natural intuition that regions which have a lot of detail are harder to compress.

8.3.4 Image Enhancement

We illustrate how base – detail decomposition can be used for image enhancement. In Fig. 8.8-bottom, we plot $B(x) + \vartheta R(x)$, for different ϑ values, where ϑ is a parameter indicating the amount of enhancement. Another example is shown in Fig. 8.9. Our approach opens the doors to segment-wise manipulation, which is useful in common situations like when segments have different illumination.

Note that the widely used bilateral filter [44] is very local compared to our segmentation-based approach. In Fig. 8.10 we show an example of base-detail decomposition produced by bilateral filtering. For example, the top-right cloud in the image cannot be separated as a detail by the bilateral filter, but it is successfully separated as detail using our approach.

Fig. 8.8 *Top*: Original $I(x)$, base $B(x)$, detail $R(x)$, detail magnitude $||R(x)||_2$ for better visualization. *Bottom*: Base + detail $B(x) + \vartheta R(x)$, with different amounts of detail, $\vartheta = \{0.5, 1, 2, 4\}$

Fig. 8.9 Example of enhanced image. Weak details can be multiplied to become more visible with respect to the base

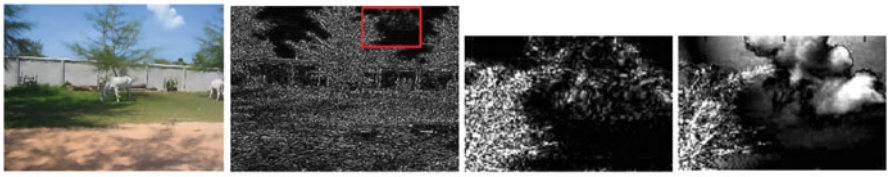

Fig. 8.10 Limitations of bilateral filter. From left to right: Bilaterally filtered (BLF) image; Residual (detail) of the bilateral filtering; Zoom-in of the residual; Zoom-in of the detail that our segmentwise base-detail decomposition produces

8.3.5 Saliency

Images of three-dimensional scenes contains structure at different scales and resolutions. Humans often need to foveate specific image locations to acquire higher level of details. For example, a small image blob might correspond to a person walking towards you and require further investigation. In this work we consider only bottom-up attention where fixation saliency is used to predict the first few seconds (3 s) of free-viewing of an image. The prediction consists of a probability map which does not take order of fixation into account. (By contrast, in top-down attention humans actively search for specific objects or scene structures.)

Our saliency model takes as input the base-detail decomposition $B_c(x), R_c(x)$, generated for a partition $p_{t_{max}}(I(x))$, defined in (8.4), whose minimum homogeneity threshold is t_{max} (Sect. 8.3.3). Note that candidate regions are not used here. Each image pixel x is assigned to the segment $i(x)$ which contains it and we define $|\mathcal{D}_{i(x)}| = size(\mathcal{D}_i)$, if $x \in \mathcal{D}_i$, where $\{\mathcal{D}_i \in p_{t_{max}}(I(x))\}$ are the segments of the partition. Similarly, we evaluate each segment's average detail and assign this value to all pixel positions of the segment support, obtaining $A(x) = \dfrac{1}{size(\mathcal{D}_i)} \displaystyle\sum_{z \in \mathcal{D}_i} R^A(z)$ if $x \in \mathcal{D}_i$. Here, $R^A(z)$ is the mean of the detail's $n_c = 3$ color channels at position

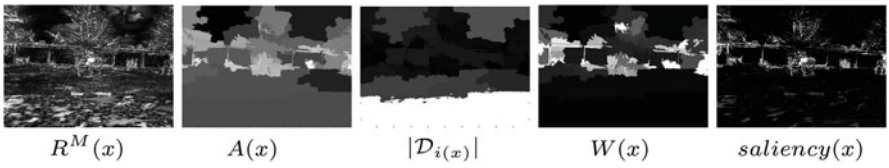

Fig. 8.11 From left to right: Maximum-channel detail $R^M(x)$; segmentwise average detail $A(x)$; segment size $|\mathcal{D}_{i(x)}|$; weight factor $W(x) = \sqrt{A(x)/|\mathcal{D}_{i(x)}|}$; $saliency(x) = R^M(x)[(1-\gamma)W(x)+\gamma]$

Fig. 8.12 An illustration of how our saliency model penalizes the detail in large roughly-homogeneous segments. The representation on the *right* is obtained by $I'_c(x) = B_c(x)+W(x)R_c(x)$, for each color channel c

z, that is, $R^A(z) = \frac{1}{n_c}\sum_{c=1}^{n_c} R_c(z)$. We use the segment sizes and the segmentwise average detail to weight the maximum-channel detail $R^M(x) = \max_{c=1}^{n_c} R_c(x)$ (see Fig. 8.11). The weight we propose is given by $W(x) = \sqrt{A(x)/|\mathcal{D}_{i(x)}|}$.

$$saliency(x) = R^M(x)\left[(1-\gamma)\,W(x)+\gamma\right]. \tag{8.7}$$

Here, γ is a small number, $\gamma = 0.15$ in our experiments. It allows to keep a fraction of $R^M(x)$ unweighted. This is useful for pixels whose weight is close to zero, $W(x) \approx 0$. The γ parameter means that the detail $R(x)$ is never completely ignored.

Intuitively, we relate the detail (Fig. 8.11-left) to bottom-up saliency. However, we penalize detail which belongs to large segments, without eliminating it completely (Fig. 8.11-right). An illustration of how an image would look like with this kind of detail penalization is shown in Fig. 8.12. The use of the segment size as an important saliency factor could be related to figure-ground pre-attentive mechanisms in V1. In terms of V1 neuron responses, very small regions tend to be highlighted against larger regions [50], but in this paper we do not address neurophysiology.

Our hypothesis is that regions which cannot be described by a simple model require foveation. This is the case of small regions with a lot of detail. The segments that are less likely to require foveation are those which are fit well by a simple

polynomial model (have little detail), as well as those which have detail but are large. In the latter case, the detail is likely to be due to a texture pattern, e.g., grass.

Classical models (e.g., [30]) use multiscale processing. Instead of this, our segment-based approach adapts to local scales of images. Also, unlike classical models, we do not explicitly use neural mechanisms such as center-surround receptive fields and lateral inhibition mechanisms. But it can be argued that base-detail decomposition is implicitly accomplishing similar functions.

The proposed fixation saliency method predicts human fixations. Note that this is different from salient object proposals. It is possible to link human fixations predictions and candidate regions by machine learning, as shown in [38]. But in this work we do not address this issue.

8.4 Experiments

In this section, we present results of the candidate region proposals (Sect. 8.3.2) and the bottom-up saliency (Sect. 8.3.5) as a prediction of free-viewing human fixations. Both of them are based on the bottom-up segmentation we propose (Sect. 8.3.1). The fundamental theory behind the saliency method is the base-detail decomposition (Sect. 8.3.3). We do not evaluate base-detail decomposition and image enhancement because there is no natural way of doing it. We do not focus on image segmentation, so we do not include experiments on it.

8.4.1 Datasets

Many of the classic datasets are biased because they were collected with a specific purpose, i.e., for saliency experiments. They are mostly composed of iconic photographs, presenting a clearly salient and centered object over a simple background. But this is highly atypical of natural images, which typically include many objects with complex relations and partial occlusions (humans rarely see iconic images). Hence it arguably more realistic to study saliency on natural image datasets such as Pascal, which has been one of the leading reference benchmark in Computer Vision for the last years. Recently, Hou et al. [38] released the free-viewing fixations of 8 participants on a subset of 850 images of Pascal (first 3 s). In this subset we have an average of 5.18 foreground objects per image and an average of 2.93 background objects. An extreme case is the rightmost image in Fig. 8.13, which has 52 foreground objects, most of which are far from the center of the image. A representative case is the third from the right image in Fig. 8.13, with 6 foreground objects.

For our candidate regions experiments, we use a subset of 1,288 images of Pascal VOC, for comparison with [6], as detailed in [10]. For the bottom-up saliency experiments, we use the 850 images of Pascal-S which include human fixations. We

Fig. 8.13 (**a**) Examples of iconic images from ImgSal [36]; (**b**) Examples from a non-iconic dataset: Pascal VOC [17]

also experiment on the 1,003 images of the standard dataset Judd [31], which can be considered non-iconic, although we have no statistics of the number of objects or their distribution in the images.

8.4.2 Candidate Regions

In this section, we evaluate the coverage of our candidate regions. Initially, we obtained an average of 116 selected-segments per image after selection and from these we make an average of 721 combinations which constitute the pool of candidate regions per each image. The evaluation metric is Intersection over Union (IoU), which accounts the number of pixels of the intersection between a candidate region and a groundtruth region, divided by the number of pixels of their union.

We evaluate the generated candidate regions with the 57-classes ground truth, containing both foreground and "stuff" classes. We compare our Candidate Regions (CR) to three state-of-the-art methods. (I) The classical Constrained Parametric Min-Cuts (CPMC) [15] method is designed for foreground objects, which explains its better performance on foreground objects. Their overall performance on the 57 classes is lower than our performance. (II) In [6], the segment combinations are generated by taking combinations of the 150 segments (on average) that their hierarchical segmentation approach outputs for each image. Their method is more sophisticated than ours and we observe that they tend to get larger and less homogeneous segments than we do. Our performance is lower but comparable: 74 % IoU versus 77 % for [6]. But we achieve it with nearly half the number of combinations – 721 compared to 1,322 – and with a simpler and faster algorithm (4 s per image in its Matlab prototype). In Table 8.1 we refer to their segments as UCM-combs and to our candidate regions as CR-combs. (III) The Selective Search [46] method is competitive in terms of speed. Our method outperforms theirs on the region candidates task (74.0 % compared to 67.8 % IoU), with less than half the number of proposals. (Note, however, that [46] present results for bounding boxes and not for regions). See an example of the proposals generated by our CR-combs method in Fig. 8.14. See Table 8.1 with the region-based IoU and recall results.

Table 8.1 Region-based IoU (in %) comparison. CPMC [15], UCM [6], Sel. Search [46], and our CR – candidate regions. Boldface denotes the first and second best results

	All IoU	Recall (%)	# cands.	Time (s)
CPMC	59.6	57.6	**150**	250
UCM-combs	**77.0**	**80.0**	1,322	850
Sel. search	67.8	66.1	2,100	**4**
Our CR combs	**74.0**	**70.3**	**721**	**4**

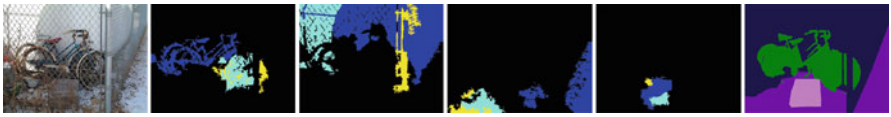

Fig. 8.14 Left-to-right and top-to-bottom: Original image, top three segments for bike, wall, snow, rock, and ground truth. Note that the segments are good even for object classes that perform poorly overall (e.g., bike)

8.4.3 Saliency

The fixation saliency method that arises from our unified approach predicts free-viewing human fixations surprisingly well. Despite only accounting for saliency within segments and not taking into account inter-segment saliency, our method is among the highest ones in complex datasets like Pascal-S [38] and Judd [31]. Pascal is a particularly interesting case because state-of-the-art fixations methods have low performance on it (perhaps because they were developed and tested for iconic images). In Pascal our method outperforms the state of the art. On the Judd dataset only AWS [21] outperforms our method.

In Fig. 8.15 we show a comparison of our Base-Detail Saliency (BDS) method, Adaptive Whitening Saliency, AWS [21], Image Signature, SIG [28], and L. Itti's original model [30]. In Fig. 8.16 we show some examples for qualitative comparison between the results of the different algorithms.

It is hard to determine the failure modes of our saliency algorithm. Our reliance on segmentation may seem problematic. It is known that segmentation is an ill-posed problem and no low-level segmentation algorithm exists that can reliably detect the boundaries of objects without top-down assistance. But our approach is more robust because we rely only on a proto-segmentation. Still errors in the segmentation can cause errors in the base-detail decomposition which may cause our approach to fail.

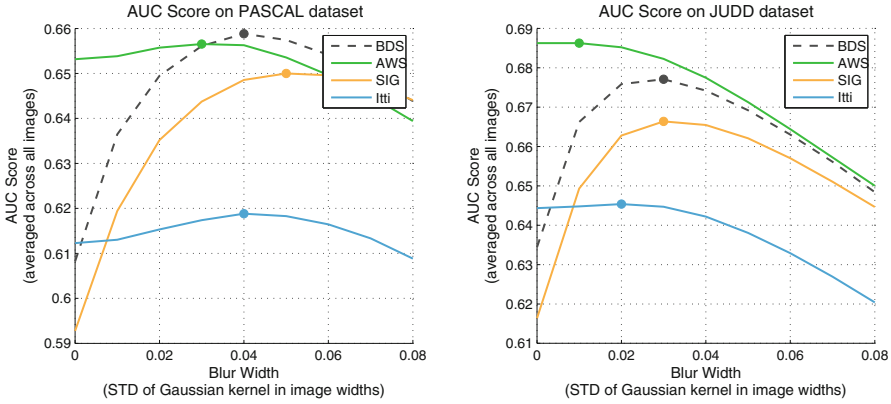

Fig. 8.15 Bottom-up saliency performance. *Left*: Pascal-S dataset [38]. *Right*: Judd dataset [31]. Approaches compared: Our Base-Detail Saliency (BDS), Adaptive Whitening Saliency (AWS, [21]), Image Signature (SIG, [28]), L. Itti's original model (Itti, [30])

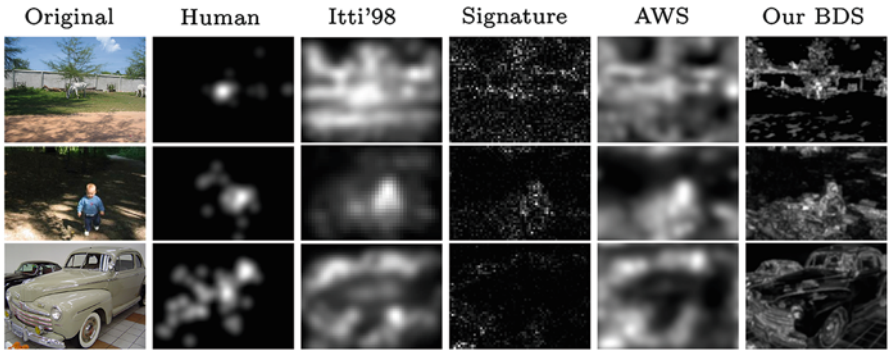

Fig. 8.16 From left ro tight: Original; Human – fixations collected on 8 subjects with free-viewing task, first 3 s [38]; Itti's original model [30]; Spectral signature [28]; AWS [21]; Our Base-Detail Saliency (BDS) Bonev and Yuille

8.5 Conclusions

We propose a unified approach addressing a set of early-vision bottom-up processes: segmentation, candidate regions, base-detail decomposition, image enhancement, and saliency for fixations prediction.

Our unified approach allows the segmentwise decomposition of the image into "base" and "detail". This proves to be more versatile than a local smoothing of the image. It provides directly for image enhancement, for a novel model of fixation saliency. It is related to other vision topics which are usually formalized as different problems.

We show state-of-the-art results on our candidate regions and on our saliency for free-viewing fixation prediction. For the latter we use the psychophysics data available for the Pascal VOC dataset, which is non-iconic and particularly difficult for the state-of-the-art saliency algorithms.

Acknowledgements We would like to thank Laurent Itti, Li Zhaoping, John Flynn, and the reviewers for their valuable comments. This work is partially supported by NSF award CCF-1317376, by ONR N00014-12-1-0883 and by NVidia Corp.

References

1. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. TPAMI 34(11):2274–2282
2. Alexe B, Deselaers T, Ferrari V (2012) Measuring the objectness of image windows. TPAMI 34(11):2189–2202
3. Alpert S, Galun M, Brandt A, Basri R (2012) Image segmentation by probabilistic bottom-up aggregation and cue integration. TPAMI 34(2):315–327
4. Arbelaez P (2006) Boundary extraction in natural images using ultrametric contour maps. In: Proceedings of the 2006 conference on computer vision and pattern recognition workshop, CVPRW '06. IEEE Computer Society, Washington, DC, pp 182–
5. Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. TPAMI 33(5):898–916
6. Arbelaez P, Hariharan B, Gu C, Gupta S, Malik J (2012) Semantic segmentation using regions and parts. In: CVPR, Providence
7. Bae S, Paris S, Durand F (2006) Two-scale tone management for photographic look. ACM Trans Graph 25(3):637–645
8. Barron JT, Malik J (2012) Color constancy, intrinsic images, and shape estimation. In: ECCV, Florence
9. Barrow HG, Tenenbaum JM (1978) Recovering intrinsic scene characteristics from images. Technical report 157, AI Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025
10. Bonev B, Yuille AL (2014) A fast and simple algorithm for producing candidate regions. In: European conference on computer vision (ECCV 2014), Zurich
11. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. IEEE Trans Pattern Anal Mach Intell 35(1):185–207
12. Borji A, Sihite DN, Itti L (2013) Objects do not predict fixations better than early saliency: a re-analysis of Einhäuser et al.'s data. J Vis 13(10):18
13. Borji A, Cheng M, Jiang H, Li J (2014) Salient object detection: a survey. CoRR, abs/1411.5878
14. Bradley C, Abrams J, Geisler WS (2014) Retina-v1 model of detectability across the visual field. J Vis 14(12):22
15. Carreira J, Sminchisescu C (2012) CPMC: automatic object segmentation using constrained parametric min-cuts. TPAMI 34(7):1312–1328
16. Einhäuser W, Spain M, Perona P (2008) Objects predict fixations better than early saliency. J Vis 8(14):18
17. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A (2010) The PASCAL visual object classes (VOC) challenge. Int J Comput Vis 88(2):303–338
18. Farbman Z, Fattal R, Lischinski D, Szeliski R (2008) Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Trans Graph 27(3):67:1–67:10

19. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. IJCV 59(2):167–181
20. Galun M, Sharon E, Basri R, Brandt A (2003) Texture segmentation by multiscale aggregation of filter responses and shape elements. In: ICCV '03, Nice, pp 716–
21. Garcia-Diaz A, Leborán V, Fdez-Vidal XR, Pardo XM (2012) On the relationship between optical variability, visual saliency, and eye fixations: a computational approach. J Vis 12(6):1–22
22. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6(6):721–741
23. Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. Neuron 65(2):150–164
24. Golub GH, Van Loan CF (2012) Matrix computations, vol 3. JHU Press, Baltimore
25. Gonzalez RC, Woods RE, Eddins SL (2004) Digital image processing using matlab. Pearson Prentice Hall, Upper Saddle River
26. Gorelick L, Basri R (2009) Shape based detection and top-down delineation using image segments. Int J Comput Vis 83(3):211–232
27. Horn BKP, Brooks MJ (1986) The variational approach to shape from shading. Comput Vis Graph Image Process 33(2):174–208
28. Hou X, Harel J, Koch C (2012) Image signature: highlighting sparse salient regions. IEEE TPAMI 34(1):194–201
29. Humayun A, Li F, Rehg JM (2014) RIGOR: reusing inference in graph cuts for generating object regions. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), Columbus. IEEE, New York
30. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE TPAMI 20(11):1254–1259
31. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: ICCV, Kyoto, pp 2106–2113. IEEE, New York
32. Land EH (1977) The retinex theory of color vision. Sci Am 237(6):108–28
33. Leclerc YG (1989) Image and boundary segmentation via minimal-length encoding on the connection machine. In: Proceedings of a workshop on image understanding workshop, Palo Alto. Morgan Kaufmann, San Francisco, pp 1056–1069. ISBN 1-55860-070-1. http://dl.acm.org/citation.cfm?id=94703.99744
34. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324
35. Leonenko N, Pronzato L, Savani V (2008) A class of Rényi information estimators for multidimensional densities. Ann Statist 36(5):2153–2182
36. Li J, Levine M, An X, He H (2011) Saliency detection based on frequency and spatial domain analyses. In: Proceedings of BMVC, Dundee, pp 86.1–86.11. http://dx.doi.org/10.5244/C.25.86
37. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans Pattern Anal Mach Intell 35(4):996–1010
38. Li Y, Hou X, Koch C, Rehg JM, Yuille AL (2014) The secrets of salient object segmentation. In: CVPR, Columbus
39. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Henry Holt and Co., New York
40. Mottaghi R, Chen X, Liu X, Fidler S, Urtasun R, Yuille A (2014) The role of context for object detection and semantic segmentation in the wild. In: CVPR, Columbus
41. Russ JC, Woods RP (1995) The image processing handbook. J Comput Assist Tomogr 19(6):979–981
42. Shapley R, Enroth-Cugell C (1984) Visual adaptation and retinal gain controls. Prog Retin Res 3:263–346
43. Todorovic S, Ahuja N (2008) Region-based hierarchical image matching. IJCV 78(1):47–66
44. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Sixth international conference on computer vision, 1998. IEEE, Washington, DC, pp 839–846

45. Tu Z, Zhu S-C, Shum H-Y (2001) Image segmentation by data driven Markov chain Monte Carlo. In: Proceedings of eighth IEEE international conference on computer vision, 2001. ICCV 2001, Vancouver, vol 2, pp 131–138
46. Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM (2013) Selective search for object recognition. Int J Comput Vis 104(2):154–171
47. Woodham RJ (1980) Photometric method for determining surface orientation from multiple images. Opt Eng 19(1):191139–191139
48. Xu C, Xiong C, Corso JJ (2012) Streaming hierarchical video segmentation. In: ECCV, Florence
49. Yuan L, Sun J (2012) Automatic exposure correction of consumer photographs. In: Fitzgibbon AW, Lazebnik S, Perona P, Sato Y, Schmid C (eds) ECCV (4). Volume 7575 of Lecture notes in computer science. Springer, Berlin/New York, pp 771–785
50. Zhaoping L (2003) V1 mechanisms and some figure-ground and border effects. J Physiol 97(1):503–515
51. Zhaoping L (2014) Understanding vision: theory, models, and data. Oxford University Press, Oxford
52. Zhu SC, Yuille A (1996) Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans Pattern Anal Mach Intell 18(9):884–900
53. Zhu L, Chen Y, Lin Y, Lin C, Yuille A (2012) Recursive segmentation and recognition templates for image parsing. IEEE Trans Pattern Anal Mach Intell 34(2):359–371
54. Zhu Y, Zhang Y, Yuille A (2014) Single image super-resolution using deformable patches. In: 2014 IEEE conference on computer vision and pattern recognition (CVPR), Columbus, pp 2917–2924
55. Zitnick CL, Dollár P (2014) Edge boxes: locating object proposals from edges. In: ECCV, Zurich

Chapter 9
Perception-Based Motion Cueing: A Cybernetics Approach to Motion Simulation

Paolo Pretto, Joost Venrooij, Alessandro Nesti, and Heinrich H. Bülthoff

Abstract The goal of vehicle motion simulation is the realistic reproduction of the perception a human observer would have inside the moving vehicle by providing realistic motion cues inside a motion simulator. Motion cueing algorithms play a central role in this process by converting the desired vehicle motion into simulator input commands with maximal perceptual fidelity, while remaining within the limited workspace of the motion simulator. By understanding how the one's own body motion through the environment is transduced into neural information by the visual, vestibular and somatosensory systems and how this information is processed in order to create a whole percept of self-motion we can qualify the perceptual fidelity of the simulation. In this chapter, we address how a deep understanding of the functional principles underlying self-motion perception can be exploited to develop new motion cueing algorithms and, in turn, how motion simulation can increase our understanding of the brain's perceptual processes. We propose a perception-based motion cueing algorithm that relies on knowledge about human self-motion perception and uses it to calculate the vehicle motion percept, i.e. how the motion of a vehicle is perceived by a human observer. The calculation is possible through the use of a self-motion perception model, which simulate the brain's motion perception processes. The goal of the perception-based algorithm is then to reproduce the simulator motion that minimizes the difference between the vehicle's desired percept and the actual simulator percept, i.e. the "perceptual error". Finally, we describe the first experimental validation of the new motion cueing algorithm and shown that an improvement in the current standards of motion cueing is possible.

Keywords Motion cueing • Motion perception • Self-motion • Simulation • Model predictive control • Washout

When we move through the environment, our central nervous system (CNS) is called upon to create a continuous estimate of the state of our own body with respect to the world (i.e. its position, orientation and their derivatives). This perceptual process,

P. Pretto • J. Venrooij • A. Nesti • H.H. Bülthoff (✉)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
e-mail: heinrich.buelthoff@tuebingen.mpg.de

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_9

generally referred to as self-motion perception, plays a crucial role in vital tasks such as balance and locomotion. Not surprisingly, considerable research efforts have been devoted to further our understanding of how the CNS integrates available sensory information to create an internal representation of the physical bodily motion [7, 21]. These studies have provided great insight into how physical stimuli are transduced into neural information by the visual, vestibular and somatosensory systems and how this information is processed in order to create a percept of self-motion. From experiments on human perception, theories and models are developed and continuously improved, which have as ultimate goal to provide accurate descriptions of the perception of the complex motion patterns experienced in everyday life.

A widely employed approach to the study of self-motion perception is the so called Cybernetics approach. The term "Cybernetics" generally refers to the study of those systems able to generate changes in their environment which, in turn, are fed back to the system, altering its status and influencing future actions [39]. When applied to perception, the Cybernetics approach considers the brain of biological organisms as a complex control system where subcomponents can be isolated and individually investigated. Experimental methods such as psychophysics are employed to quantify the properties of human self-motion perception. Modelling methods rooted in system theory are used to model how the brain infers a representation of the physical world from sensory signals and generates actions to successfully interact with it. In this perspective, the Cybernetics approach, the psychophysical methods and behavioral measurements described in this chapter are powerful tools for Cognitive Engineering to investigate the fundamentals of perceptual and cognitive processes. The use of simulation technologies enables the implementation of these fundamental processes in control algorithms and improves the current understanding of human perception and action.

The study of self-motion perception is of great benefit for a wide variety of fundamental but also applied fields. For example, it is of use for the development of perceptual tests for clinical diagnosis and rehabilitation of patients with balance disorders [1, 20]. In this chapter, we focus on a different, but equally practical application: we will address how a deep understanding of the functional principles underlying self-motion perception can be exploited to develop new algorithms for motion simulators.

Motion simulators are widely employed in many different applications, such as training, research and development and entertainment. Despite large differences in architecture, complexity and purpose, all motion simulators have one aspect in common: the use of a Motion Cueing Algorithm (MCA), also known as motion drive algorithm. The MCA is responsible for converting a desired physical motion into commands that are sent to the motion simulator. In this chapter, we will explore how the design of MCAs can benefit from fundamental knowledge on self-motion perception and, in turn, how motion simulation can increase our understanding of brain perceptual processes.

9.1 Self-Motion Perception and Vehicle Simulation

With the progress in science and technology, the use of simulations of human controlled tasks became more and more common. This not only thanks to the increased fidelity of such simulations, but also to the effectiveness of training, studying, entertaining, or otherwise involving humans in simulations. A common application of simulation is vehicle simulation, where the manual control of a vehicle such as an aircraft, car or ship is simulated. Such simulations typically include control devices similar to the control devices in the actual vehicle and a visual projection from the point of view of the driver or pilot of the simulated vehicle. In some cases, the simulation includes motion cues, provided to the human operator through a motion platform (such as a hexapod motion simulator). The goal of including motion cues is to increase realism of the simulation by providing additional cues to the human operator. This type of vehicle simulation is referred to as *motion-based simulation* or, in short, motion simulation.

In motion simulation, any reproduction of a simulated maneuver on a motion simulator is always constrained by the physical limits of the motion system. For example, maneuvers that involve high sustained accelerations, such as the take-off of an airplane, make the simulator reach its physical limits eventually. At this point, the pilot inevitably experiences a conflict between the expected and experienced motion. This has several potential drawbacks such as the occurrence of motion sickness, an overall degrade in simulation realism and the occurrence of "false training", i.e., when pilots learn to react in a way that is not appropriate for the simulated maneuver.

As the quality of a motion simulation is eventually determined by the simulator user, based on how well the desired perception of motion is reproduced, it should be evident that knowledge on self-motion perception is beneficial, if not a prerequisite, in the development and evaluation of MCAs. However, there is a surprisingly large gap between what we know of human self-motion perception and how much of that knowledge is actually used in practice. In Sect. 9.3, we will explore how we can bridge this gap by using perception-based motion cueing. However, before doing so we need to describe more in detail the traditional approach to the motion cueing challenge.

9.2 The Motion Cueing Challenge

Technological advances in the field of vehicle simulation have mainly concerned the development of vehicle models, the visual rendering, control loading and quality of the auditory stimuli. Motion cueing, on the other hand, remains one aspect of motion simulation that has not benefited equally from technological advances, and still remains one of the main challenges when operating a motion simulator. In

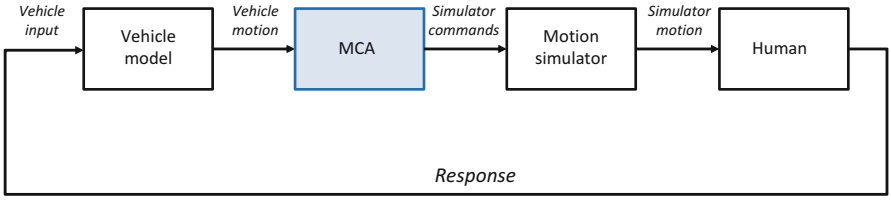

Fig. 9.1 The process of motion simulation. Vehicle motion provided by a model is converted into simulator commands by the MCA. The human (driver) experiences the simulator motion and reacts on the vehicle commands, changing the state of the vehicle model

fact, many state-of-the-art simulators are still employing MCAs that are not so different from those developed several decades ago, for simulators whose hardware and software are, today, obsolete.

Motion cueing can be defined as the conversion from desired motions to motion simulator input commands. For example, if one wants to simulate a car driving on a race track in a motion simulator, the output from a vehicle model can be used to compute the desired motion, expressed in, e.g., the accelerations and rotational rates of the car. An MCA processes this desired motion and calculates the appropriate motion simulator input commands – expressed in, e.g., position and velocity of the motion simulator's actuators (see Fig. 9.1). For instance, in case the simulation is executed on a hexapod simulator, where the actuators are six telescopic jacks, the MCA determines the input commands for each actuator at every simulation time step. From the above description it becomes apparent that an MCA needs to be tailored to the configuration and capabilities of the motion simulator hardware (e.g., number and configuration of its actuators). Next to that, the MCA is tailored to the vehicle that is simulated, the maneuver that the vehicle is performing and possibly even the person driving the simulated vehicle.

As the range of motions of a simulated vehicle, such as the car driving on the race track, is typically much larger than the motion range (or workspace) a motion simulator can cover, the MCA needs to ensure that its output commands are realizable by the platform. In order to provide a realistic simulation, even for maneuvers that are outside the motion simulator's workspace, several different approaches can be applied. In literature, different MCAs have been proposed, tested and compared. It is not the purpose of this chapter to provide an exhaustive overview of all of these algorithms, or even of the most important ones (see e.g. [8] for a concise overview). Instead, in the following we will sketch – in broad strokes – how motion cueing algorithms were initially conceived and what their limitations are. This helps to understand the context of the alternative approach to motion cueing that we will address in the next section.

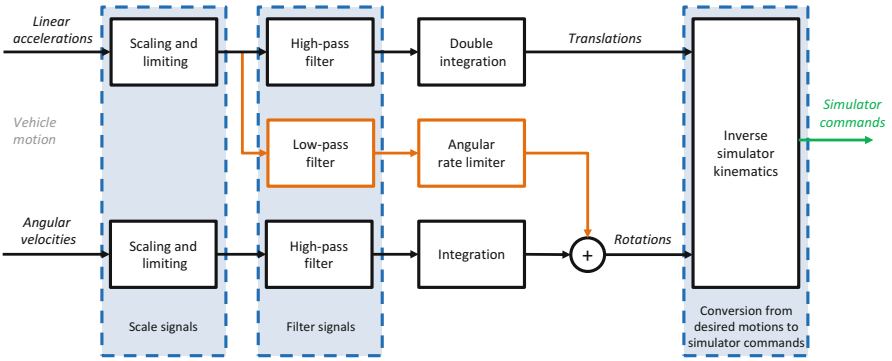

Fig. 9.2 Schematic representation of a CMCA

9.2.1 Classic Approach

The first MCA described in literature is what has become known as the "classical" algorithm. This approach was first described by Conrad and Schmidt [5] and later extended by Reid and Nahon [29–31]. In this chapter, this approach will be referred to as the Classical MCA (CMCA). The workings of the CMCA are illustrated in Fig. 9.2.

The desired motion is expressed in terms of linear accelerations and angular velocities. Both are (possibly) scaled and limited through saturation to account for the physical limits of the simulator. The results of this process are then high-pass filtered, which allows for the onsets of motions (high frequency signal) to be reproduced whilst removing the sustained accelerations and angular velocities, likely irreproducible within the simulator workspace. The output of the high-pass filter can be integrated to obtained, e.g., desired simulator position and simulator orientation. The application of filters (both high-pass and low-pass filters, as described below) is the hallmark of the "classical" approach.

In order to reproduce also the sustained low-frequency components of the acceleration, most CMCA use a procedure called tilt-coordination. Tilt-coordination exploits the fact that human acceleration sensors such as the otolith organs respond to the vector sum of gravity and inertial accelerations (i.e. the gravito-inertial vector). This can lead to an ambiguity between a static head tilt and a linear acceleration with the head upright [2]. Through tilt-coordination, the simulator cabin is tilted to align gravity with the gravito-inertial vector resulting from sustained accelerations. For example, when the body is tilted backwards the vestibular system receives inputs that are indistinguishable from when the body is accelerated forwards. This ambiguity is used in tilt-coordination to reproduce sustained longitudinal and lateral linear accelerations by pitch and roll rotations, respectively. To ensure that the perceptual ambiguity is resolved in favor of the perception of a translation rather

than a physical tilt, visual cues congruent with linear self-motion are concurrently provided [11] and the tilt rate of the simulator cabin is maintained below the human perceptual threshold [13, 27, 36, 40]. The latter is done by limiting the rate at which the platform tilts, using an angular rate limiter. The tilt-coordination path of the classic MCA is indicated in Fig. 9.2 ("Angular rate limiter").

Next to the high-pass filters and the tilt-coordination, many CMCAs have an additional feature that continuously repositions the simulator towards the center of its operational workspace, i.e., towards the position where the simulator dexterity is highest. This is achieved by designing the high-pass filters such that the simulator always returns to its initial position. With these filter settings all simulator motions are 'washed out' over time. For this reason, CMCA are also referred to as washout filters or washout MCAs.

9.2.2 Limitations of Classic Approach and Evolution of MPC-Based Motion Cueing

Alternative approaches to the CMCA described above have been proposed in literature. Many of these, however, are variations and extensions of the CMCA approach, sharing therefore most of its problems and limitations. Among these, the most important are:

- Limited consideration for perceptual factors: Although several studies have shown that the human sensitivity to motion is influenced by factors such as stimulus characteristics and cognitive factors [23, 27, 33], these and many other insights into human self-motion perception are not included in currently used MCAs. A common issue in motion simulation that would immediately benefit from an increased consideration for perceptual factors is, for example, the relatively high rate of simulator sickness, cause by sensory conflicts [14].
- Need for extensive tuning: an MCA has a number of parameters that can be set to a desired value. Typically, the number of parameters is in the order of several tens but can sometimes be as high as a few hundreds. With these parameters, the behavior of the MCA can be adapted to better fit the needs of the simulator user, simulator architecture, simulated maneuver, etc. However, determining the best value for all parameters is a difficult job, making it an expensive and time consuming process, which can only be done by experts. Many simulator users do not have the expertise to improve the tuning of their MCA or adapt it for different maneuvers.
- Limited use of simulator motion envelope: the tuning of an MCA is typically done using a "worst-case" scenario, such that the worst (largest) expected motion still fits within the motion space of the motion simulator. This implies that the simulator capabilities are not fully exploited during normal operation.

The abovementioned problems are aggravated by the fact that no systematic method exists to evaluate the quality of an MCA. Many simulator users (e.g., airline pilot training centers, car manufacturers, research institutes, universities) would like to improve their MCAs but have no tools at their disposal to measure their quality. Even experts tuners have to rely on subjective judgments, for example from test drivers, in order to adjust the tuning. An objective, reliable and repeatable method of evaluating the quality of an MCA would be welcomed by the motion simulation community at large. In Sect. 9.4 a methodology that was developed by the authors to fill this gap is presented.

Over the years, several filter-based variations of the classical (washout) approach were implemented to improve MCAs by addressing one or more of the above problems. For example, in adaptive washout filters the parameters are tuned automatically through an optimization algorithm. Although this reduces the need for parameter tuning, it introduces a new problem, namely the definition of a cost function. The optimal solution is found by minimizing the value of this cost function [28]. A cost function also contains parameters, such as weights of cost elements, which require tuning. In the adaptive washout approach the problem of tuning is not removed, but merely relocated.

More recently, it was proposed to solve the motion cueing challenge using Model Predictive Control (MPC) [4, 6]. MPC is a process control strategy that relies on dynamical models of the process. The application of MPC to a driving simulation scenario requires models of the motion simulator and of the driver's perception. Simulator commands are then obtained by minimizing the perceptual error (mismatch) without exceeding the physical limitations of the simulator. The MPC approach optimizes at each time step the simulator commands for a finite time-horizon based on a prediction of the future. The commands obtained for the current time step (i.e., the first value of the control sequence) are used for controlling the simulator and the process is repeated at each subsequent time step. Note that, for offline optimization, the entire trajectory is known and there is therefore no need to develop and implement a predictor, nor to use a prediction horizon shorter than the duration of the entire trajectory. The benefits of MPC motion cueing are that the platform motion matches the vehicle motion for as long as possible and that simulator limits are explicitly taken into account, thus eliminating the need to tune the MCA for the worst-case motion. In the published implementations of MCA algorithms for motion simulation, the focus has been centered on obtaining a real-time implementation. For this reason, linear simulator models were implemented in MPC algorithms to reduce computational load. Similarly, linear models of human sensory dynamics were usually favored over nonlinear self-motion perception models. However, the MPC approach also allows for implementation of nonlinear models which better describe human self-motion perception and motion simulators with nonlinear dynamics.

9.3 Perception-Based Motion Cueing

The perception-based motion cueing (PBMC) approach is based on advanced MPC control strategies for nonlinear models, and differs from the traditional approach CMCA in several ways. The most important difference is that PBMC aims to reproduce the *perception* of motion, instead of reproducing physical motion. Another important difference is that PBMC operates through optimizing simulator input commands, instead of filtering the motion that is to be reproduced. This approach is being developed with the purpose of improving the control strategies of the simulator on one side, and the perceptual fidelity, realism and user experience on the other side.

9.3.1 PBMC Structure

Knowledge about human self-motion perception is included in the MCA and used to calculate how the motion of a vehicle is perceived by a human observer inside a simulator. The calculation is possible through the use of a self-motion perception model, which simulate the brain's motion perception processes. The model transforms the linear and rotational components of vehicle inertial motion into a corresponding "vehicle percept", i.e. the mental representation of vehicle motion in a "perceptual space" (Fig. 9.3). It is then possible to map any vehicle motion within the physical space into its corresponding vehicle percept in the perceptual space. Thanks to a simulator model, the same calculation can be done for any simulator motion, which has a corresponding simulator percept. It is therefore possible to predict how a certain motion (trajectory) in the actual simulator workspace would "feel" in its corresponding perceptual space. The goal of the PBMC approach is then to minimize the difference (mismatch) between the vehicle desired percept and the actual simulator percept, i.e. the "perceptual error" (Fig. 9.3). The PBMC approach makes use of an iterative optimization process to find the simulator motion that best approximates the perception one would have in the actual vehicle. This optimization process searches for the minimum cost of a so called "perception-based cost function".

In the PBMC approach, the self-motion perception model, simulator model, perceptual cost function and optimization algorithm are integrated in a software framework. In the followings a brief description of these components and their integration is provided.

9.3.1.1 Self-Motion Perception Model

A human self-motion perception model (in the following also referred to as perception model) is a computational model that aims to describe the continuous

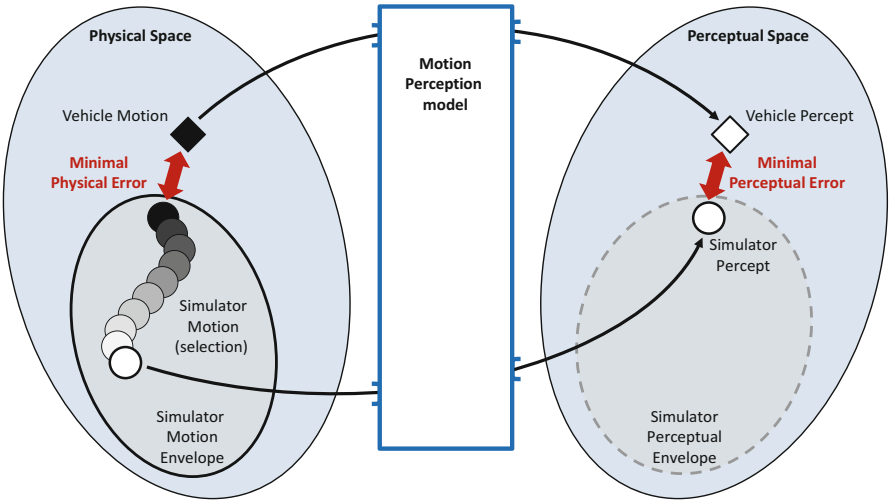

Fig. 9.3 Concept of the perception-based motion cueing

dynamical process taking place in the brain to produce a perception of motion through the environment. It does this by reproducing the way our sensory organs retrieve relevant information from the environment, and how these sensory signals are combined in the brain.

Physical motion is sensed by humans for a large part through the vestibular system, which is comprised of two components: the semicircular canals and the otoliths. These sensors respectively sense rotational motions (i.e., rotational velocities) and linear accelerations (i.e., specific forces). In addition, the eyes receive visual information (i.e., optical flow). The perception model has, therefore, two classes of inputs: *vestibular inputs*, i.e., rotational velocities and specific forces and *visual inputs*, i.e., motion information obtained through the visual sensors.

The first processing step of the perception model consists of reproducing the *sensory dynamics* of the perceptual organs. The next step constitutes the *sensory integration*, in which the different sensory information is combined into a single percept of motion. The third (optional) step accounts for the relative perceptual sensitivity to different stimuli, by transforming the physical units into perceptual units. This is done by applying psychophysical *perceptual laws*, which are nonlinear analytical functions that relate the intensity of the sensory input to the corresponding perceived intensity of motion [35]. This final step allows for the comparison of motion intensities within different degrees of freedom (e.g. rotations vs. translations) and accounts for nonlinearities in perception that are not related to the properties of the physical stimulation. For example, it is known that, for increasing motion intensities, the human sensitivity decreases in a nonlinear fashion [19, 22, 23]. The output of the perception model consists of a multi-dimensional *percept*: perceived motion described in perceptual units.

A perception model – and knowledge on human motion perception in general – can be beneficial for motion-based (vehicle) simulation applications in several different ways. For example, with a perception model one can determine how a given motion trajectory in a given simulator with a given MCA is perceived by the person inside the simulator. This gives indications with regards to whether the trajectory is reproduced realistically, which simulator provides a superior simulation and/or how the MCA – and/or the motion simulator – can be improved to increase the realism of the simulation.

9.3.1.2 Simulator Model

In PBMC, the simulator model serves two purposes: first of all, it is used to compute the simulator motion, i.e., the motion in response to simulator input commands. The accuracy of the simulator kinematics is essential, as the simulator motion is used to calculate the simulator percept, i.e., the perception that results from the given simulator motion.

The second purpose of the simulator model is to provide information regarding workspace constraints, in other words: which dynamical states the simulator can reach. Information on the state and workspace constraints of the simulator is absent in many traditional motion cueing approaches. This can, and often does, result in a situation where motion commands, sent to the simulator by the MCA, would cause one of the simulator's actuators to reach a position or velocity limit. Using a simulator model in the motion cueing approach avoids such issues. In fact, it allows for a more optimal use of the workspace that is available, as the simulator's capabilities are known at any given time.

9.3.1.3 Perception-Based Cost Function

The perception model provides two percepts: the simulator percept and the vehicle percept, which are both multidimensional quantities. The PBMC approach aims at minimizing the difference between these two percepts. The cost function is essential in achieving that goal. The cost function provides a measure for differences across the dimensions and weighs their relative contribution. Minimizing the cost function corresponds to minimizing discrepancies between vehicle and simulator motion (the "perceptual error" from Fig. 9.3). This provides answers to questions such as how much linear acceleration can be represented by tilting (tilt-coordination) or how much the simulator can be moved without the driver noticing the motion (sub-threshold motions). Some traditional MCAs intend to incorporate some knowledge on human self-motion perception, for example, by limiting the maximum allowable tilt-rate. However, research has shown that sensitivities to motions vary with factors such as the presence of visual information, motion complexity, and/or active vehicle control [24, 27]. A perception-based cost function allows for implementing these scientific findings on human self-motion perception in a more systematic and extensive way.

9.3.1.4 Optimization Algorithm

The purpose of the optimization algorithm is to find the simulator input commands for which the cost function is minimized within the given constrains. In other words: it finds the allowable simulator input whose percept best approximates the percept in the actual vehicle. For PBMC, several different optimization algorithms can be employed, as long as they are capable of working with constrained nonlinear multivariable functions.

9.3.1.5 Perception-Based Motion Cueing Framework

By bringing all the elements discussed above together, the PBMC framework is obtained. The flow diagram in Fig. 9.4 provides a simplified illustration of this framework. A vehicle motion is fed to the perception model to compute the vehicle percept, which describes the perception one would have inside the vehicle. The challenge is now to reproduce this same percept in the motion simulator, using appropriate simulator commands. These are calculated using a model of the motion simulator. The simulator motion is passed through the perception model, providing the simulator percept. Note that the perception models computing the simulator and vehicle percept are identical. The cost function provides a quantitative measure of similarity between the two percepts (i.e., the cost), which is used by the optimization algorithm to compute a new set of simulator commands with the aim to decrease the cost. This optimization process is repeated until the cost is reduced below a suitable tolerance level. The optimized simulator commands that result from this

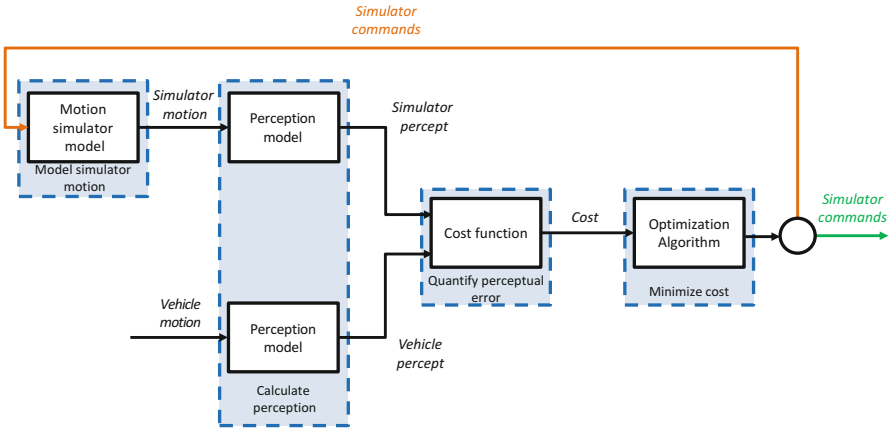

Fig. 9.4 Schematic representation of the PBMC framework. The goal of the PBMC approach is to minimize the cost, i.e., the difference between the simulator percept and the vehicle percept. The optimization of the simulator commands is done iteratively by an optimization algorithm until the optimized simulator commands are obtained

optimization are fed to the actual simulator. Given that models are accurate, the cost function is well-defined and the optimization algorithm successfully minimizes the cost, it is guaranteed that the optimized simulator commands provide a simulator percept that best approximates the vehicle percept.

9.3.2 Applications, Benefits and Limitations

Perception-based motion cueing can increase the realism of motion simulation mainly because the algorithm has more information available than most alternative algorithms. This additional information largely resides in the perception model – providing information on how vestibular and visual stimuli are perceived and integrated into a percept of motion – and the simulator model – providing information on the state and constraints of the simulator.

The perception model allows for exploiting the limitations and ambiguities of human perception. For example, as described in Sect. 9.2.1, under certain circumstances linear acceleration and static tilt are indistinguishable from each other. A perception model, combined with a perception-based cost function, is able to identify and make use of these circumstances. Although this "perceptual trick" appears in many traditional MCA implementations, PBMC does not require the typical tuning and is capable of dynamically adapting the tilt-translation ratio over the course of the simulation. Another example is the exploitation of absolute and differential perceptual thresholds: the human sensory system can only perceive motion stimuli of certain intensity. Stimuli smaller than this absolute threshold are not perceived. Similarly, the human sensory system can only distinguish stimuli when the difference between them is larger than a certain differential threshold. Both types of threshold values depend on many different factors (e.g., motion direction, frequency, intensity, etc.). These factors vary over time, and thus so do the threshold values. By implementing knowledge on these varying thresholds, PBMC is able to determine which motions can and cannot be perceived and which motions can and cannot be distinguished.

The simulator model allows for making more optimal use of the simulator's available workspace. Using the simulator model, PBMC explicitly accounts for actuator limits, exploiting therefore the full spectrum of simulator capabilities. Unlike traditional approaches, PBMC does not scale down the physical motion so that even the most aggressive part of a maneuver still fits within the simulator's workspace. Instead, PBMC processes unscaled physical motion and finds, by means of a perception-based cost function, the simulator inputs that result in the best available cueing, based on the simulator's capabilities. PBMC provides cueing that approaches, but never actually reaches, the actuator limits, making optimal use of the available hardware.

The inclusion of a simulator model makes PBMC also useful in studying simulator concepts themselves. For example, PBMC can be employed in cost-benefit analyses for simulator upgrades, such as the addition of a new axis or the

improvement of an existing one. Similarly, it can provide important insight into the development of a simulator concept that does not (yet) exist. PBMC automatically provides the optimal cueing without a dependency on parameters such as the washout filter gains and filter frequencies and is therefore better suited than many traditional methods for the above mentioned scenarios.

The main disadvantage of the PBMC over CMCA is the relative complexity of the algorithm, which inevitably impacts on the speed at which the algorithm operates. At the time of writing, real-time simulations have not yet been achieved. Hence, all tests with PBMC have been performed on pre-recorded maneuvers and offline execution of the PBMC algorithm. For many applications, online calculation operating in real-time are required. In order to achieve this, the current implementation of the PBMC algorithm needs to be further optimized for speed performance. Another drawback that derives from the relative complexity of PBMC is the increased effort to implement a PBMC algorithm. However, this is of less concern, as it can be largely addressed through the development of user-friendly interface solutions. Another disadvantage of the PBMC algorithm here presented is that, in the attempt to bring target and simulator percept as close as possible to each other, it will inevitably favor an inhomogeneous scaling of the trajectory over its time evolution. For instance, in the cueing of a maneuver consisting of two consecutive decelerations, where the first deceleration is twice as strong as the second, the 2:1 proportionality is easily lost if the first deceleration cannot be faithfully reproduced by the simulator. In contrast, in a traditional MCA, a scaling factor is tuned for the entire maneuver to ensure that it fits into the simulator capabilities, and proportionality is therefore more likely maintained.

9.4 Validation of the Perception-Based Motion Cueing Approach

The PBMC approach here introduced would be little more than a conceptually interesting idea, with little lasting impact or practical application, if it would remain without a rigorous and objective evaluation. The goal of such an evaluation is to show the potential benefits and improvements with respect to the current motion cueing approaches. In order to make this comparison possible, a novel methodology was developed, based on psychophysical methods that are typical of human perception research.

9.4.1 The Choice of a Suitable Method

Evaluating the quality of perceived motion is a very complex task, requiring a subjective judgment that can be affected by several factors, like the sensitivity of different sensory modalities, memory, personal preferences, experience and familiarity with the judgment task. Therefore, it is common practice to evaluate

the quality of motion cueing by using questionnaires and experts' interviews (which can be executed in more or less rigorous fashion), accompanied by various types of rating scales. Oft-used rating scales are the Visual Analog Scale (VAS) [38], where the response is specified by indicating a position on a continuous line between two end-points, and the Likert-type rating scales [16], where the response is specified by choosing a judgment from a symmetric agree-disagree scale for a series of statements (typically 5, 7, or 9 statements, ranging from completely negative to completely positive judgment). However, both questionnaires and rating scales have several disadvantages that could affect the experimental validation of motion cueing approaches.

Questionnaires and interviews with experts allow for a large degree of freedom in the response, providing typically only qualitative indications. As a result, verbal reports are not a systematic, repeatable and reliable measure of subjective judgments, and should be used only in exploratory studies. Rating scales, on the other hand, provide quantitative numerical estimates and can refer to subtle dimensions or aspects that contribute to the overall judgment. These methods are therefore better suitable for more rigorous statistical analysis. However, also rating scales have features that are disadvantageous for our current purposes, notably the resolution of the Likert-type scale, the presence of scale boundaries and the challenge of qualifying the numerical results with meaningful verbal descriptors. Likert scales have the disadvantage of low resolution (only a few possible responses), which also lead respondents to quickly reach the scale boundaries (once the maximum or minimum rating is given, there is no possibility to assign even higher or lower ratings). It is a general weakness of discrete rating systems that two trials perceived as different could, due to the lack of rating options, be assigned to the same verbal descriptor. This may result in insufficient sensitivity for measuring the multifaceted quality of motion cueing. The VASs, thanks to their analogue nature overcome this limitation of the Likert scales. Nevertheless, it remains difficult to determine appropriate verbal descriptors for the subjective responses along the scale and to quantify their relative distances, as they might not be linearly paced. This limitation becomes even more evident considering that the subjective interpretation of verbal labels may differ among individuals and cultures. In general, rating scales suffer from the limitation of collecting subjective responses that are at best on an ordinal scale, and providing data that are treated as on an interval/ratio scale [10].

These considerations motivated a search for an alternative method providing the following features: (i) sufficient sensitivity to distinguish among subtle differences in the perceive quality; (ii) accounting for subjective differences in the meaning of verbal and numerical labels; (iii) providing numerical quantification on a ratio scale, which allows for traditional statistical analysis and (iv) providing unambiguous interpretation of distances between the judgment scores. We found that the method of *magnitude estimation* with *cross-modality matching* paradigm was suitable for our purposes. This method was originally introduced by Stevens to measure the perceived magnitude of physical intensities [34]. Later the same method was successfully adopted to investigate perceptual aspects of non-physical dimensions,

like political opinions [17, 18], aesthetic preferences [12] and linguistic judgments [3]. To the best of our knowledge, the study described here utilizes the magnitude estimation method in the field of motion perception and simulation research for the first time. With this method it is possible to express a judgment about certain features of a stimulus, providing an estimate in any response modality. At the foundations of this method lie the following two assumptions: firstly, the human observer constructs an abstract internal representation of the stimulus' features, which is independent of the sensory modality used to retrieve this representation. Secondly, the observer implicitly attributes a magnitude to the stimulus features that can be expressed on a ratio scale [34]. In an actual experiment, these assumptions allow an observer to produce the quality judgments in any response modality, based on the relative difference in the magnitude of the examined feature over multiple trials. In the study presented here, two response modalities were used to express the magnitude of perceived quality of motion cueing: numerical estimate and line production. Producing a numerical estimate consists of providing a number; while hand-drawing a horizontal line constitutes a line production task. Using magnitude estimates from two response modalities enables us to test the internal validity of the measurement scale: if the participants in the experiment are indeed rating the motion quality on a ratio scale, this should result in consistent answers across the two different modalities. For a detailed description of the cross-modality matching as a subjective assessment technique, the reader is referred to [26]. The method was used for different purposes throughout the different experimental phases, as described in the experimental procedure below.

9.4.2 Validation Experiment

The following MCAs were tested and compared over three car maneuvers:

- *Classic (CLA)*: classic washout algorithm (cf. Fig. 9.2) using the same filter parameters for all maneuvers
- *Classic Tuned (CLT)*: classic washout algorithm, using filter parameters specifically tuned for each maneuver to better exploit the motion envelope of the simulator
- *Non-weighted Perception-Based (NPB)*: perception-based algorithm using a non-weighted sum of linear accelerations and angular rotations in the cost-function of the optimization algorithm
- *Weighted Perception-Based (WPB)*: perception-based algorithm, similar to the previous one, but with a weighted sum of linear accelerations and angular rotations in the cost function

Eight participants (average age: 29.3 years, 1 female) were recruited through the participants' database of the Max Planck Institute for Biological Cybernetics. They declared to hold a full and valid driving license for cars and to perform active driving

on a regular basis. They had normal or corrected-to-normal vision. All participants provided informed written consent prior to their inclusion. The study was conducted in accordance to the Declaration of Helsinki (1964).

9.4.2.1 Experimental Procedure

The experimental procedure consisted of three phases, the *calibration* phase, the *experiment* phase and the *verbal qualification* phase.

In the *calibration* phase, participants were familiarized with the magnitude estimation method and the cross-modality matching task. They were asked to assign numerical estimates to different lines of varying lengths and to draw lines whose length was proportional to numbers of varying magnitude. This procedure ensured that participants were able to produce consistent magnitude estimates ($R > .95$) across different response modalities before the actual experiment.

In the *experiment* phase, participants were accommodated inside the CyberMotion Simulator (CMS) and exposed to motion trajectories (Fig. 9.5). The CMS is a dynamic simulator that was developed to expand the limited workspace and dexterity of traditional hexapod-based simulators [25]. It is an 8-degrees-of-freedom serial robot, where a 6-axes industrial robot manipulator is mounted on a linear rail and equipped with a motorized cabin at the end effector (Fig. 9.5, left). The cabin is equipped with two 1920×1200 projectors (Eyevis, Germany) and interference filter stereo projection system (Infitec GmbH, Germany), which provide up to 160×90 deg Field-of-View on the cabin inner side (Fig. 9.5, right). The cabin is also equipped with mounting possibilities for haptic control devices used for flight and driving simulation, as well as for head motion and gaze tracking.

The motion trajectories consisted of synchronized video, inertial and audio recordings of a car performing three maneuvers: strong braking, going around a roundabout and slalom. Inertial data were obtained from an inertial navigation system, while the video was captured with two full-HD cameras positioned near the driver's head in the actual car [37]. The use of recorded trajectories ensured that participants were exposed to the same motion stimuli and immersed in a realistic multi-sensory (visual, auditory and inertial) simulation. This also allowed the use of 3D video projection to further enhance the simulation realism. During the trajectory playback, two superimposed auditory signals (beeps) indicated the beginning and the end of each maneuver.

Participants were instructed to pay attention to the motion information ("*what you feel*") in between these two beeps and how it relates to the visual information ("*what you see*"). After every trajectory playback, they were required to draw a horizontal line and provide a numerical estimate. These had to be proportional to either their impression of the motion aspect that was asked for. They were asked about the following aspects[1]:

[1]For brevity, the first aspect is later referred to as "motion direction", and the second as "motion strength".

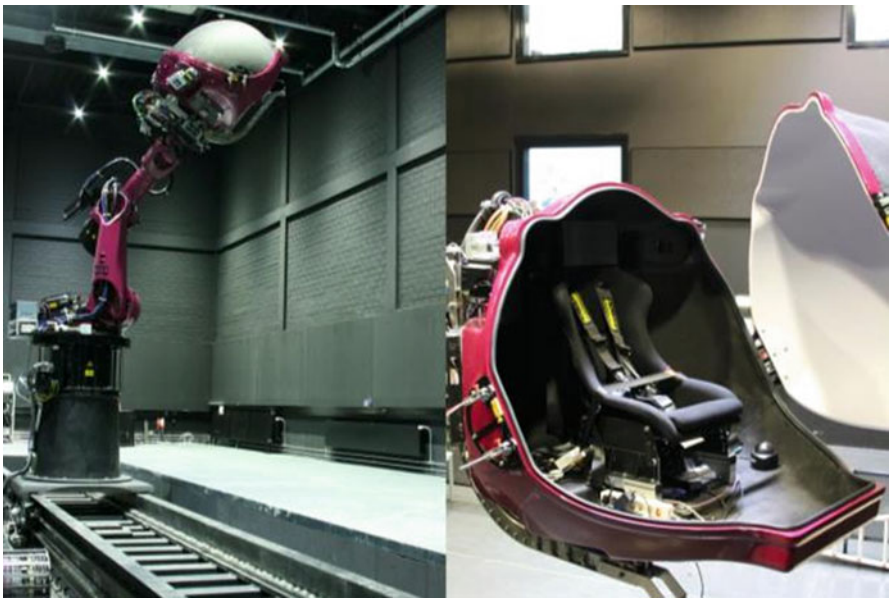

Fig. 9.5 The MPI CyberMotion Simulator. Exteriors (*left*) and cabin interiors (*right*)

- *"Agreement of motion direction: how well does the direction of the motion you feel agree with the direction of the motion you see?"* High level of agreement between the visual and inertial motion directions would result in long lines and large numbers, low level of agreement would result in shorter lines and smaller numbers.
- *"Appropriateness of motion strength: How appropriate is the strength of the motion you feel compared to the motion you see?"*Appropriate motion strength (neither too strong nor too weak) would result in long lines and large numbers, inappropriate motion strength would result in shorter lines and smaller numbers.

The questions appeared on the screen, while the simulator was steady in an upright position. The participants were instructed to provide any line length/positive number for the first answer, keeping in mind that they might want to provide longer/shorter lines and larger/smaller numbers later on. The two questions were assessed on separate blocks of trials, with the order of the questions balanced across participants. Each of the three maneuvers was reproduced with four different MCAs, for a total of 12 trajectories per block. The order of the maneuvers and the order of MCAs per maneuver were randomized across participants. A different maneuver (double lane change) was used as training before each block to familiarize the participants with the questions, the response procedure, the different MCAs and the simulation environment.

As additional data, the following information was recorded: motion sickness questionnaires were collected for all participants before and after the experiment

phase [15]. During the experiment, the level of sickness was monitored every 10 min using a numerical score [9]. After the experiment, participants filled out a questionnaire to report their subjective ratings about mental demand, level of concentration, ability to maintain a constant level of attention, level of frustration and physical comfort on a 9-point rating scale.

In the *verbal qualification* phase, participants were asked to evaluate the perceived magnitude of a set of verbal qualifiers [32] indicating "quality" (e.g., "good", "bad", "so-so") using the same method of the previous phases, i.e. magnitude estimation task with cross modality matching (numerical estimate and line production). This allows for a verbal interpretation of the numerical results of the MCAs quality ratings, and provides an indication of the subjective distance between verbal qualifiers.

Overall, the evaluation experiment lasted about 2.5 h and was structured as follows:

- Calibration task: magnitude estimation of lines and numbers
- Baseline questionnaire about motion sickness
- Training on the first question (i.e., strength or direction, 4 trials)
- First experiment block: 12 trials
- Verbal qualifiers: magnitude estimation of verbal qualifiers
- 15 min break
- Training on the second question (i.e., strength or direction, 4 trials)
- Second experiment block: 12 trials
- Verbal qualifiers: magnitude estimation of verbal qualifiers
- Post experiment questionnaire about motion sickness

9.4.2.2 Data Analysis and Results

One participant did not complete the experiment due to mild symptoms of motion sickness and was therefore excluded from the analysis. The remaining seven participants showed a high correlation ($R > .95$) between numerical estimates and lines length in the calibration phase, and were therefore included in the experiment and in the data analysis. The raw numerical estimates and lines lengths collected in the experiment were normalized and a combined standard score was computed from the mean of the normalized values. A similar procedure was adopted for the analysis of the verbal qualifiers. The results of the motion cueing quality rating are shown in Fig. 9.6, plotted against the verbal descriptors and the overall standard score.

Due to the different characteristics of the tested maneuvers, and the supposed independence of the judgments between the rated motion aspects, considerably different patterns of results were to be expected. Therefore, a repeated-measure analysis of variance (rmANOVA) was run independently for each maneuver (brake, roundabout or slalom) and rated aspect (motion direction or strength) to test the effect of MCAs on the perceived quality of motion cueing. Post-hoc tests with Bonferroni correction for multiple comparisons were used. The significant results are reported in Table 9.1.

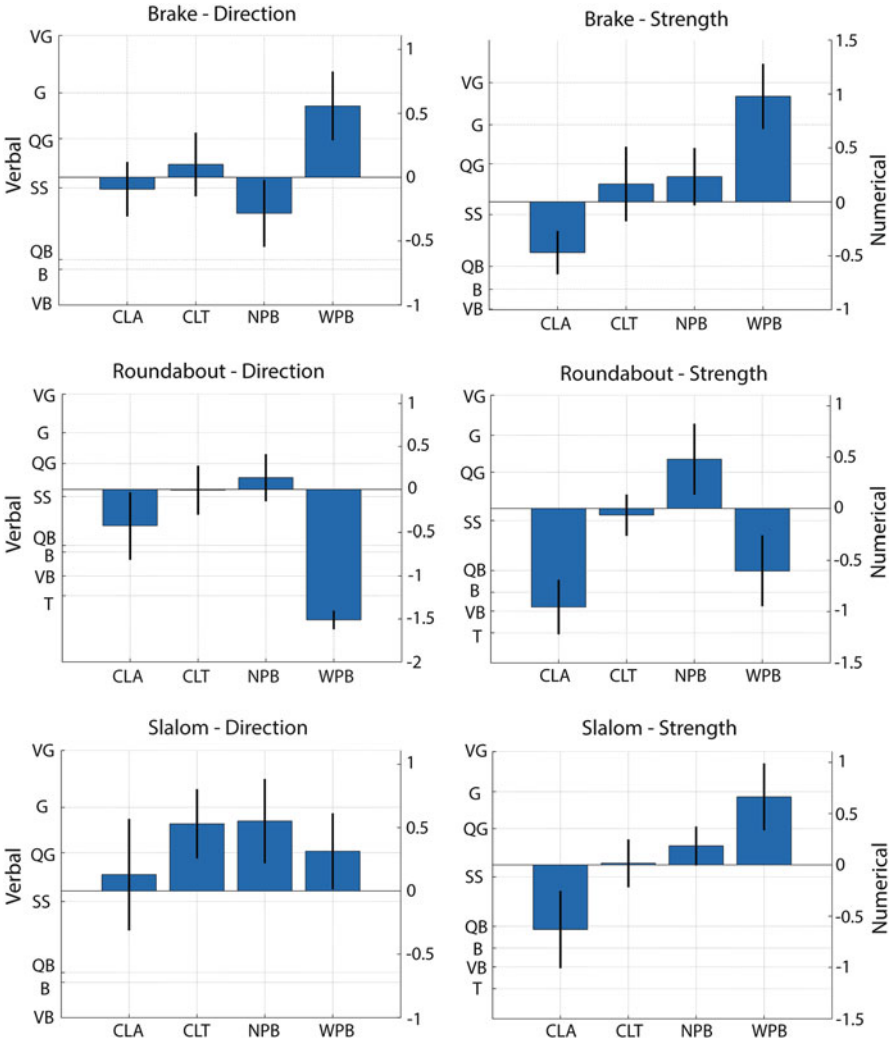

Fig. 9.6 Averaged standard score across subjects of the quality ratings of PBMC (3rd and 4th bar of each plot), as compared to CMCA (1st and 2nd bar of each plot). The *lines* indicate standard error. Each plot refers to one of the maneuvers (brake, roundabout or slalom) and rated aspect (motion direction or strength). Verbal qualifiers are indicated on the left vertical axis: *VG* very good, *G* good, *QG* quite good, *SS* so-so, *QB* quite bad, *B* bad, *VB* very bad, *T* terrible

The PBMC algorithm was rated as good as the CMCA or better in all conditions, except for the roundabout. With regards to the appropriateness of motion strength, the results show a clear preference for either version of the PBMC in all the tested maneuvers. In terms of motion direction, i.e. the level of agreement between the direction of inertial motion as compared to the direction of the visual motion, the performance of the perception-based approach is comparable to the filter-based approach in the brake and the slalom maneuvers; while it is lower in the roundabout.

Table 9.1 Significant results of rmANOVA and post-hoc tests

rmANOVA (maneuver, motion aspect)	Post-hoc tests
$F(3,28) = 4.38$, $p < 0.05$(brake, motion strength)	WPB > CLA
$F(3,28) = 6.97$, $p < 0.01$(roundabout, motion direction)	CLT > WPBNPB > WPB
$F(3,28) = 4.54$, $p < 0.05$(roundabout, motion strength)	NPB > CLA
$F(3,28) = 3.38$, $p < 0.05$(slalom, motion strength)	WPB > CLA

Nevertheless, in the roundabout maneuver the NPB algorithm receives similar quality ratings as the classic algorithms.

Overall, the results of this first validation experiment are encouraging, as they indicate that, already in its first implementation, the perception-based solution is positively rated. Moreover, the results show that the methodology chosen for the experiment is appropriate for the collection of quantitative data and the use of inferential statistics, which can be used to find differences on the quality score of MCAs.

9.5 Conclusions and Future Development

The first experimental validation of the PBMC approach described here has shown that an improvement in the current standards of motion cueing algorithms is possible. Remarkably, this approach shows great potential for future applications, as it can provide advantages to both simulator users and engineers. The quality of a classical MCA depends strongly on the tuning. Tuning is expensive and can only be done by experts. By reducing the amount of tuning required, simulator users can increase the quality of their simulation while reducing costs. Finally, the classical approach does not make optimal use of the simulator capabilities. By improving the way the MCA exploits the hardware, simulator users increase their return on investment.

As suggested by the data collected in the roundabout maneuver, more work is needed to further investigate the relation between measured overall quality and the weights given to each motion cue in the perceptual cost function. Therefore, more knowledge is needed on the relation between the measured quality of motion cueing algorithms and the quality as assumed by the optimization algorithms. With this regard, the difficulty of measuring subjective experience and preferences remains a main challenge. To gain a better understanding of this relation it is essential to know how this measured MCA quality evolves over time. Further research is ongoing to include in the evaluation methodology also continuous measurement of perceived coherence between visual and actual motion in the simulation. Due to its modular structure, PBMC will also directly benefit from advances in self-motion perception modelling, an active and dynamic research field. Among the open questions, individual differences between humans and the role of sensory and cognitive factors typical of everyday scenarios are certainly of great relevance for further development of PBMC. By continuing in its quest of filling the gap between

what is known about human self-motion perception and what is used in motion simulation, PBMC will allow for increasingly relevant improvement of the quality, realism and usefulness of motion simulations.

Acknowledgments The research described in this publication was funded by the German Federal Ministry of Education and Research under grant number 03V0138. The responsibility for the contents of this publication lies with the author. This study was supported by the Max Planck Society, by the WCU (World Class University) program and by the Brain Korea 21 PLUS Program both funded by the Ministry of Education through the National Research Foundation of Korea.

References

1. Agrawal Y, Bremova T, Kremmyda O, Strupp M, Macneilage PR (2013) Clinical testing of otolith function: perceptual thresholds and myogenic potentials. J Assoc Res Otolaryngol 14(6):905–915
2. Angelaki DE, Yakusheva TA (2009) how vestibular neurons solve the tilt/translation ambiguity. Comparison of brainstem, cerebellum, and thalamus. Ann N Y Acad Sci 1164(May):19–28
3. Bard EG, Robertson D, Sorace A (1996) Magnitude estimation of linguistic acceptability. Language 72(1):32–68
4. Beghi A, Bruschetta M, Maran F (2012) A real time implementation of MPC based Motion Cueing strategy for driving simulators. In: 2012 IEEE 51st annual conference on decision and control (CDC), Hawaii, USA, pp 6340–6345
5. Conrad B, Schmidt SF (1970) Motion drive signals for piloted flight simulators. http://ntrs.nasa.gov/search.jsp?R=19700017803
6. Dagdelen M, Reymond G, Kemeny A, Bordier M, Maïzi N (2009) Model-based predictive motion cueing strategy for vehicle driving simulators. Control Eng Pract 17(9):995–1003
7. Dichgans J, Brandt T (1978) Visual vestibular interaction: effects on self-motion perception and in postural control. In: Held R, Leibowitz H, Teuber H (eds) Perception, handbook of sensory physiology, vol 8. Springer, Berlin, pp 755–804
8. Garrett NJI, Best MC (2010) Driving simulator motion cueing algorithms – a survey of the state of the art. In: Proceedings of the 10th international symposium on Advanced Vehicle Control (AVEC), Loughborough, UK, 22nd–26th Aug, pp 183–188
9. Golding JF, Mueller AG, Gresty MA (2001) A motion sickness maximum around the 0.2 Hz frequency range of horizontal translational oscillation. Aviat Space Environ Med 72(3):188–192
10. Gould D, Kelly D, Goldstone L, Gammon J (2001) Examining the validity of pressure ulcer risk assessment scales: developing and using illustrated patient simulations to collect the data. J Clin Nurs 10(5):697–706
11. Groen EL, Bles W (2004) How to use body tilt for the simulation of linear self motion. J Vestib Res 14(5):375–385
12. Han SH, Jung ES, Jung MY, Kwak J, Park S, Choe J (1994) A psychophysical evaluation of interior design alternatives for a high speed train. Comput Ind Eng 27(1–4):397–400
13. Heerspink H, Berkouwer W, Stroosma O, van Paassen R, Mulder M, Mulder B (2005) Evaluation of vestibular thresholds for motion detection in the SIMONA research simulator. In AIAA modeling and simulation technologies conference and exhibit. American Institute of Aeronautics and Astronautics. http://arc.aiaa.org/doi/abs/10.2514/6.2005-6502
14. Johnson DM (2005). Introduction to and review of simulator sickness research [electronic resource]/David M. Johnson. Research report (U.S. Army Research Institute for the Behavioral and Social Sciences)□ 1832. Rotary-Wing Aviation Research Unit, U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Rucker
15. Kennedy RS, Lane NE, Kevin S, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3(3):203–220

16. Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22(140):1–55
17. Lodge M (1981) Magnitude scaling quantitative measurement of opinions. SAGE, Newbury Park/London
18. Lodge M, Cross DV, Tursky B, Tanenhaus J (1975) The psychophysical scaling and validation of a political support scale. Am J Polit Sci 19(4):611–649
19. Mallery RM, Olomu OU, Uchanski RM, Militchin V, Hullar TE (2010) Human discrimination of rotational velocities. Exp Brain Res 204(1):11–20
20. Merfeld DM, Priesol A, Lee D, Lewis RF (2010) Potential solutions to several vestibular challenges facing clinicians. J Vestib Res 20(1):71–77
21. Mergner T, Rosemeier T (1998) Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions—a conceptual model. Brain Res Rev 28(1–2):118–135
22. Naseri AR, Grant PR (2012) Human discrimination of translational accelerations. Exp Brain Res 218(3):455–464
23. Nesti A, Barnett-Cowan M, Macneilage PR, Bülthoff HH (2014) Human sensitivity to vertical self-motion. Exp Brain Res 232:303–314
24. Nesti A, Masone C, Barnett-Cowan M, Robuffo Giordano P, Bülthoff HH, Pretto P (2012) Roll rate thresholds and perceived realism in driving simulation. In: Driving simulation conference (2), Paris, pp 23–32
25. Nieuwenhuizen FM, Bülthoff HH (2013) The MPI CyberMotion Simulator: a novel research platform to investigate human control behavior. J Comput Sci Eng 7(2):122–131
26. Pepermans RG, Corlett EN (1983) Cross-modality matching as a subjective assessment technique. Appl Ergon 14(3):169–176
27. Pretto P, Nesti A, Nooij S, Losert M, Bülthoff HH (2014) Variable roll-rate perception in driving simulation. In Driving simulation conference (3), Paris, pp 40.1–40.7
28. Sivan R, Ish-Shalom J, Huang J-K (1982) An optimal control approach to the design of moving flight simulators. IEEE Trans Syst Man Cybern 12(6):818–827
29. Reid LD, Nahon MA (1985) Flight simulation motion-base drive algorithms: Part 1 – Developing and testing the equations. UTIAS report, no. 296. University of Toronto, Institute for Aerospace Studies
30. Reid LD, Nahon MA (1986) Flight simulation motion-base drive algorithms: Part 2, Selecting the system parameters. UTIAS report, no. 307. University of Toronto, Institute for Aerospace Studies
31. Reid LD, Nahon MA 12. Flight simulation motion-base drive algorithms: Part 3 – Pilot evaluations. UTIAS report, no. 319. University of Toronto, Institute for Aerospace Studies
32. Rohrmann B (2007) Verbal qualifiers for rating scales: sociolinguistic considerations and psychometric data. Project report. University of Melbourne, Melbourne
33. Soyka F, Robuffo Giordano P, Beykirch K, Bülthoff HH (2011) Predicting direction detection thresholds for arbitrary translational acceleration profiles in the horizontal plane. Exp Brain Res 209(1):95–107
34. Stevens SS (1956) The direct estimation of sensory magnitudes: loudness. Am J Psychol 69(1):1–25
35. Stevens SS (1957) On the psychophysical law. Psychol Rev 64(3):153–181
36. Valko Y, Lewis RF, Priesol AJ, Merfeld DM (2012) Vestibular labyrinth contributions to human whole-body motion discrimination. J Neurosci 32(39):13537–13542
37. Venrooij J (2014) VIA-sync: a novel vehicle motion recording platform for synchronised visuo-inertial playback. Presented at the vehicle dynamics conference 2014, Stuttgart, June
38. Wewers ME, Lowe NK (1990) A critical review of visual analogue scales in the measurement of clinical phenomena. Res Nurs Health 13(4):227–236
39. Wiener N (1965) Cybernetics or control and communication in the animal and the machine. MIT Press, Cambridge, Massachusetts
40. Zaichik LE, Rodchenko VV, Rufov IV, Yashin YP, White AD (1999) Acceleration perception. In Collection of technical papers, edited by AIAA modeling and simulation technologies conference

Chapter 10
The Other-Race Effect Revisited: No Effect for Faces Varying in Race Only

Isabelle Bülthoff, Regine G.M. Armann, Ryo Kyung Lee, and Heinrich H. Bülthoff

Abstract The other-race effect refers to the observation that we perform better in tasks involving faces of our own race compared to faces of a race we are not familiar with. This is especially interesting as from a biological perspective, the category "race" does in fact not exist (Cosmides L, Tooby J, Krurzban R, Trends Cogn Sci 7(4):173–179, 2003); visually, however, we do group the people around us into such categories. Usually, the other-race effect is investigated in memory tasks where observers have to learn and subsequently recognize faces of individuals of different races (Meissner CA, Brigham JC, Psychol Public Policy Law 7(1):3–35, 2001) but it has also been demonstrated in perceptual tasks where observers compare one face to another on a screen (Walker PM, Tanaka J, Perception 32(9):1117–1125, 2003). In all tasks (and primarily for technical reasons) the test faces differ in race *and* identity. To broaden our general understanding of the effect that the race of a face has on the observer, in the present study, we investigated whether an other-race effect is also observed when participants are confronted with faces that differ only in ethnicity but not in identity. To that end, using Asian and Caucasian faces and a morph algorithm (Blanz V, Vetter T, A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques – SIGGRAPH'99, pp 187–194, 1999), we manipulated each original Asian or Caucasian face to generate face "race morphs" that shared the same identity but whose race appearance was manipulated stepwise toward the other ethnicity. We presented each Asian or Caucasian face pair (original face and

I. Bülthoff (✉) • H.H. Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: Isabelle.buelthoff@tuebingen.mpg.de

R.G.M. Armann
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Department of Psychology, University of York, York, UK

R.K. Lee
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_10

a race morph) to Asian (South Korea) and Caucasian (Germany) participants who had to judge which face in each pair looked “more Asian” or “more Caucasian”. In both groups, participants did not perform better for same-race pairs than for other-race pairs. These results point to the importance of identity information for the occurrence of an other-race effect.

Keywords Human face recognition • Other-race effect • Race vs identity information

10.1 Introduction

10.1.1 Identity Versus Other Face Properties

Faces shown as static images differ from each other in terms of their intrinsic properties like identity, sex, eye color, and age among many others. They can be further categorized in terms of non-rigid deformations (e.g. their facial expressions, open or closed mouth, etc.), while external variations of viewing conditions allow yet another way to differentiate between them. For example, faces can differ in terms of their orientation with regard to the camera (profile or side view) or the illumination (from above, from the side).

Although computationally images of different faces viewed under the same viewing conditions are more similar to each other than images of the same faces viewed under different viewing conditions or with different expressions [1], for the human visual system this only occurs when dealing with images of unfamiliar people. For familiar faces, the visual system seems to discard spontaneously those external variations; it becomes ‘blind’ for most of the identity-independent variance and correctly recognizes the same person across a huge range of image and instance variability.

These findings demonstrate that identity-related information (idiosyncratic information) is the most important property of a familiar face and many studies have shown that we process familiar faces at this level before processing their category affiliation, e.g. whether the face is male or female (sex category), or Asian or Caucasian (race category but see [2]). In contrast, for unfamiliar faces, idiosyncratic information is a less robust factor for grouping same-identity faces together in the presence of non-rigid deformation and/or external variations [3] and there might be differences in grouping various renditions of the same person’s face with better performance for same-race faces (e.g. [4]).

10.1.2 The Other-Race Effect and Other Category Effects

It has been shown that for unknown faces the familiarity with the category they belong to (e.g. own-race or other-race) influences how we process them and how

well we can recognize them later. Many studies have demonstrated that we are better at recognizing faces belonging to categories which are more familiar to us. This is the case for the very well-known other-race effect (also called cross-race effect or own-race advantage), which describes the fact that we recognize faces of our own ethnicity better than other-race faces (for a recent review see [5]). The influence of familiarity has been reported for other categories, principally for age and sex. Same-age faces (for the observer) are better recognized than much younger or older ones [6, 7], and other studies even report a same-sex advantage [8–10]. Among these category effects, the other-race effect seems the most robust and the most abundantly studied, yet, the mechanisms at its base are still discussed [11].

10.1.3 Identity and Face Categories

Many studies involving faces of different categories, for example investigations of facial emotions, are based on stimuli which show the same identity exhibiting different facial expressions [12], thus there is no confound of changing identity information. So far, controlling identity while investigating the effect of only sex or race changes on face recognition has not been done (but see studies investigating sex alone [13, 14]). Most studies investigating the other-race effect or more generally the effect of race on face recognition have used stimuli displaying faces that differ not only in terms of their category affiliation but also in terms of their idiosyncratic features (among many other studies, see [15, 16]), leaving open the possibility that those effects might be somewhat dependent on the use of face stimuli of different identities.

10.1.4 Our Study

Greater expertise with own- than other-race faces is believed to be one of the explanations for the other-race effect as expertise gained from perceptual learning allows to better discriminate between members of a category after training [17]. Since ground-breaking advances in computer graphics and computer vision (for example [18]), it is possible to manipulate faces along high-level dimensions like sex, race and even perceived attractiveness or memorability while avoiding large changes at the level of idiosyncratic facial information. In the present study, we parametrically modified the race of individual faces without changing their identity to create “race morphs”. In other words, we targeted specifically race-related facial appearance and modified faces along that dimension only, leaving identity-related information mostly untouched. Our goal was to investigate whether stronger expertise with own-race faces would allow better discrimination between own-race faces and their race morphs than between other-race faces and their race morphs.

In two race comparison tasks, we asked the same participants to compare the race of two faces and to indicate which face was more Asian or more Caucasian looking. In our design, one face was always the original face (either Caucasian or Asian) and the other one of its race morphs. This comparison task is derived from the "better likeness task" designed by Beale and Keil ([19], see also [20]). In a 'parallel' task, both faces were presented simultaneously on the computer screen, thus memory load was reduced to a minimum in this task, while in a 'sequential' task, both images were shown sequentially, thus a small memory load was present.

10.2 Method

10.2.1 Participants

We tested 26 German participants (13 females, age: 21–63) at the Max Planck Institute for Biological Cybernetics in Tübingen, and 26 Korean participants (13 females, age: 18–30) at Korea University in Seoul. In accordance with the Declaration of Helsinki, the procedures were approved by local IRBs and signed consent forms were obtained from individual participants before the experiment.

Tübingen and Seoul participants answered orally a few questions prior testing to ensure that no participants had intensive contacts to individuals of the other race. All participants were paid volunteers and were naïve as to the purpose of the experiment.

10.2.2 Stimuli

We selected the faces of 20 Asian individuals (10 females) and 20 Caucasian (12 females) from our face database (http://faces.kyb.tuebingen.mpg.de, [18]). All faces were devoid of hair, facial hair, makeup, glasses or jewelry. Same-identity race morphs were created by morphing all original faces of one race category stepwise toward the other race while keeping their idiosyncratic features. We give here only a short description of the morphing method used to that end as it has been described in more detail elsewhere [18, 21, 22]. In our face database, all faces are in dense point-to-point correspondence. Using the morphable model of Blanz and Vetter [18] we created an Asian average face derived from all Asian faces and a Caucasian average face derived from all Caucasian faces. A race vector was then computed as the difference between both average faces for all points of the faces. We manipulated ethnicity by applying this race vector in full or in part to each original face to create a series of race morphs.

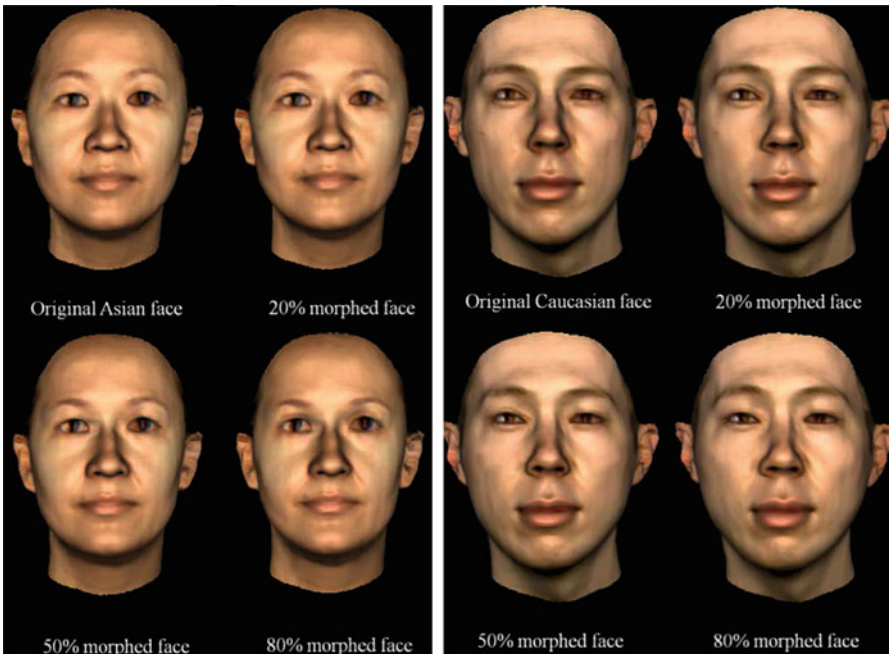

Fig. 10.1 *Left panel*: Example of an Asian original face and its morphs. *Right panel*: Example of a Caucasian original face and its morphs

Test images of all original faces and their race morphs were generated with all faces shown in frontal view. Each original face was morphed at three different levels toward the other ethnicity: 20, 50 and 80 %; with 0 % corresponding to the original face and 100 % meaning that the full ethnicity vector had been applied. Figure 10.1 shows examples of original faces of both ethnicities with their race morphs. All face images were 900-by-900 pixels in size and the approximate size of the face in the image was 600-by-730 pixels. 120 pairs were created by pairing each original face with each of its morphs. Four mask images were created by randomly mixing arbitrary face parts presented upside-down.

In each trial of the parallel task, the test faces were shown side by side on a computer screen, with the right edge of the left image and left edge of right image cut off by 50 pixels each, yielding a combined test image of 1700-by-900 pixels. For each pair, two test images were created to counterbalance left/right presentation location. In trials of the sequential task, the faces of each pair were shown alone, one after the other in the middle of the screen, with their presentation order counterbalanced across participants. The face size was the same as in the parallel task.

10.2.3 Design and Procedure

In each task participants viewed two faces in a trial. For half of the participants the question asked was "which face is more Asian?" and for the other half: "which face is more Caucasian?". Race of participants (Seoul/Tübingen) was a between-subject independent variable and race of face stimuli (Asian/Caucasian origin) and race dissimilarity (the difference in morph levels between the faces of the test pair in a trial: 20, 50, and 80 %) were within-subject independent variables. The dependent variables were accuracy and response times (RT) from correct trials. Task order was randomized across participants.

All pictures were presented in eight-bit color format on a 24-inch flat monitor screen of a Tobii T 60 XL eye tracker with 1920×1200 pixels screen resolution (refresh rate 60 Hz). Each face covered approximately an angular size of around 17 deg $\times$ 14 deg. Participants were seated in front of the computer monitor at a distance of approximately 60 cm to 70 cm.

The experiment was designed with Tobii Studio 3.1.2 software and run with a special monitor that allowed to record participants' eye gaze position in addition to their response. The eye tracking data are described elsewhere [23, 24].

In the parallel task, a trial started with a 500 ms fixation cross replaced by a response-terminated display of one face pair. Participants pressed one of two keys to indicate their choice of the left or right image. A blank screen appeared for 1 s before the start of the next trial. Presentation order of all face pairs was randomized across participants. In each trial, both face images originated from the same identity and one of them was always the original face. Every possible pair (three for each identity) was shown once. Left or right placement of the original image was randomized across trials and for each pair it was counterbalanced across participants. Though there was no time limit, participants were asked to answer as correctly and as quickly as possible. The RT in this experiment was measured from the onset of the face images to the subject's key press.

In the sequential task, each trial started with a fixation cross for 500 ms, followed by the first face for 2500 ms and a mask for 500 ms. The mask image was then replaced by a response-terminated display of the second face. Participants pressed one of two keys to indicate their choice of the first or second image. All other details are the same as in the parallel task. RT in the sequential experiment was measured from the onset of the second image to the subject's key press.

Recognition performance was measured as percent correct response (accuracy) and correct response times (RT) measured as explained above for each test condition. The RT and accuracy data were submitted to a 2 (group: Seoul vs Tübingen) $\times$ 2 (face race: Asian vs Caucasian) $\times$ 3 (pair dissimilarity: 20 % vs 50 % vs 80 %) mixed repeated-measures ANOVA, with face race and pair dissimilarity as within-participant factors and group as a between-participant factor. Statistical significance was set at $\alpha = .05$. We report partial eta-squared ($\eta_p{}^2$) as an index of effect size, with values of 0.01, 0.06, and 0.14 representing small, medium, and large effect sizes, respectively (Cohen, 1988).

10.3 Results

Because of time constraints, two of the 26 German participants and one of the Korean participants performed only one of both tasks. In addition, the data of one German participant in the sequential task was removed because he had inverted the response buttons in that task (indicated by an excessive error rate). We compared accuracy for both questions (i.e. depending on the question asked: "which one is more Asian?" versus "which one is more Caucasian?") at each race dissimilarity level. For Seoul participants, no differences in accuracy were found across both questions for the parallel task and the sequential tasks (independent t-tests: all $ps > .20$, all $ps > .19$ respectively). Similarly, no differences were found for the Tübingen participants (independent t tests: all $ps > .19$, all $ps > .20$, respectively). Therefore we collapsed all results across questions for all analyses. In agreement with the difference in memory load, participant overall performance was better in the parallel task (80 %) than in the sequential task (73 %). Owing to the nature of the tasks, response times cannot be compared.

In both tasks, for the smallest race dissimilarity level (20 %), that is when the original face and the morphed face were quite similar to each other, both Tübingen and Seoul participants performed better than chance (50 %) for both types of face stimuli (one sample t-tests, all $ps < .001$), while for the easiest condition (80 % difference) they performed significantly worse than perfect performance (all $ps < .001$). Thus the tasks did not appear to be either too hard or too easy.

10.3.1 Parallel Task

For the parallel task, the analyses show that Seoul and Tübingen participants did not differ from each other in terms of their accuracy performance ($F < 1$, ns). There was a main effect of *pair dissimilarity*, as participants' performance increased with larger race difference between the faces to compare ($F(2,96) = 138.56$, $p < .001$, $\eta_p^2 = .833$). There was also a main effect of *face race*, with better performance with Caucasian faces than with Asian faces ($F(1,48) = 6.81$, $p = 0.012$, $\eta_p^2 = .124$). These results were qualified by an interaction between *face race* and *participant group* ($F(1,48) = 8.23$, $p = 0.006$, $\eta_p^2 = .146$). There was no other significant interaction (all $ps > .115$). Separate ANOVAs for each group of participants showed a significant main effect of face race for the Seoul group ($F(1,26) = 17.52$, $p < .001$, $\eta_p^2 = 0.10$) as participants performed better with Caucasian faces ($M = 83.9$ %) than Asian faces ($M = 77.7$ %,), while Tübingen participants performed equally well with Asian faces ($M = 81$ %) and Caucasian faces ($M = 81$ %, $F < 1$, ns). Figure 10.2 shows the detailed accuracy results for both groups of participants.

Depending on conditions, participants needed on average between 2.5 and 7 s to enter their answer. The RT data were subjected to a mixed repeated-measures ANOVA involving *participant group* as between-subjects variable and

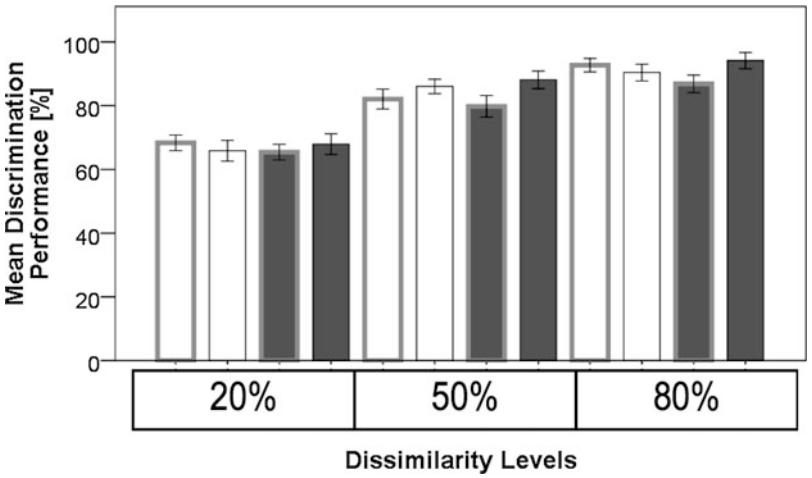

Fig. 10.2 Mean accuracy for the parallel task for Seoul and Tübingen participants as a function of race dissimilarity between test face pairs (20, 50 and 80 %) and face race (Asian or Caucasian). Each dissimilarity level is shown four times. *White bars* represent data for participants of Tübingen, grey bars for Seoul participants. Bars with a *black outline* depict the results for the Caucasian pairs, bars with a *thick grey* outline depict the results for the Asian pairs. Error bars represent SE

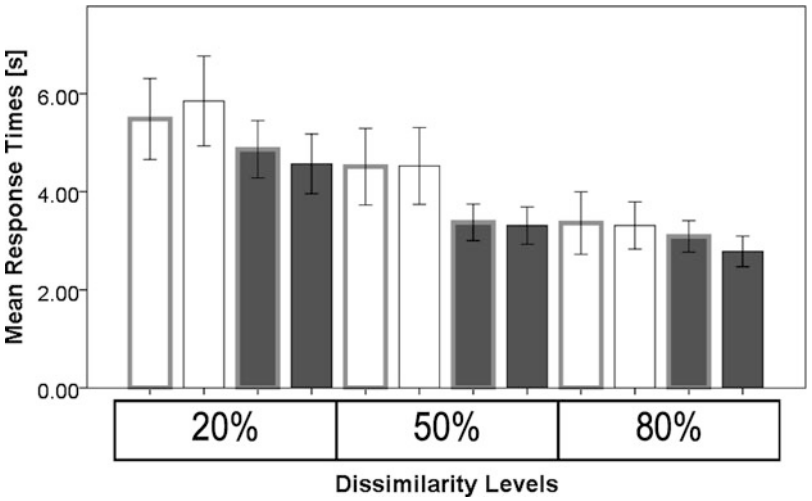

Fig. 10.3 Mean correct response times for the parallel task. For full legend, see Fig. 10.2

pair dissimilarity and *face race* as within-subjects variables. The analysis shows that only the factor *pair dissimilarity* had a significant impact on RT ($F(2,96) = 40.55$, $p < 0.001$, $\eta_p^2 = .458$). Figure 10.3 shows that participants responded more rapidly when the task was easier (i.e. for larger dissimilarity between the faces to compare). Although Tübingen participants ($M = 4.6$ s, $SE = 0.54$) were slower than Seoul

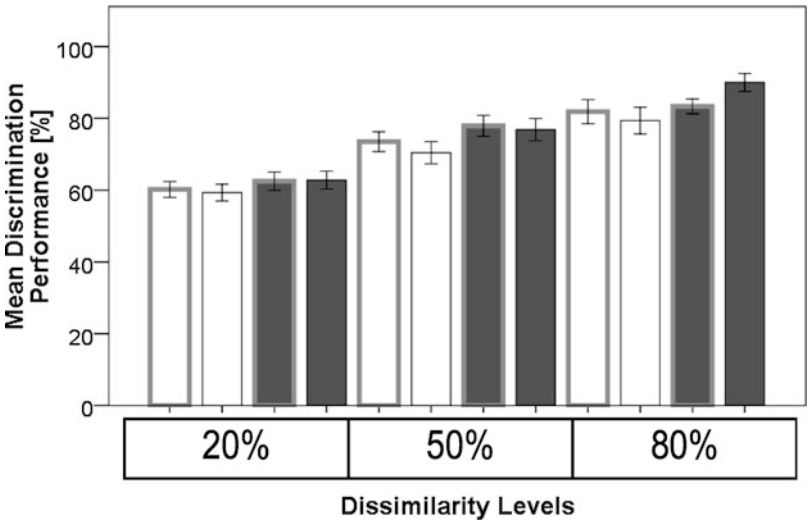

Fig. 10.4 Mean accuracy for the sequential task. For full legend, see Fig. 10.2

participants (M = 3.4 s, SE = 0.53) the difference was not significant because of the large variability of RT responses. There was also no significant main effect for *participant group* or *face race* (all Fs < 1, ns) and no significant interaction (all *p*s > .129). These results indicate that the factor *pair dissimilarity* affected response times to same-race and other-race face pairs for both groups similarly.

10.3.2 Sequential Task

When the faces were shown sequentially, Seoul participants were better overall than Tübingen participants, but this difference did not reach significance ($F(1,48) = 3.53$, $p = 0.066$, $\eta_p^2 = 0.068$). The factor *face race* did not affect response accuracy (F < 1, ns), so that the only significant main effect was for the *pair dissimilarity*, ($F(2,96) = 112.25$, $p < 0.001$, η_p^2 .700). As the faces got more dissimilar, accuracy improved. There was no significant interaction (all *p*s > .089). Figure 10.4 shows the details of the accuracy data.

Depending on condition, participants needed on average between 1.6 and 2.2 s to answer. Figure 10.5 shows that participants responded more rapidly as the dissimilarity increased ($F(2,96) = 9.58$, $p < 0.001$, $\eta_p^2 = .166$), no other significant main effects or interactions were found (all *p*s > 0.072). These results indicate that the factor dissimilarity affected response times to same- and other-race face pairs for both groups similarly.

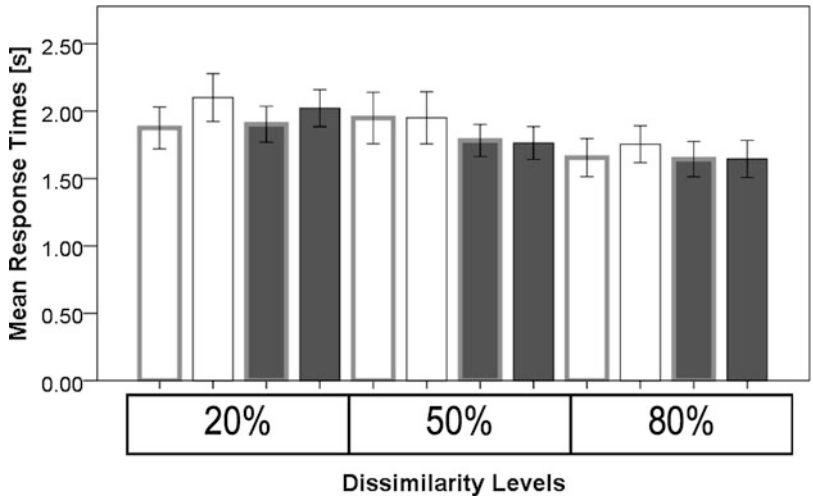

Fig. 10.5 Mean correct response times for the sequential experiment. For full legend, see Fig. 10.2

10.4 Discussion

In this study, we investigated whether participants showed better performance for same-race than other-race faces in a race comparison task. Although the ethnicity of the faces used as stimuli have been shown to influence participants' response in many previous studies (for example, among many others: [11, 25]; for a review see [26]), most of the faces used in these earlier studies displayed concomitant changes of ethnicity and identity. In the present work we specifically avoided any changes in identity when manipulating face race. We tested participants in a parallel and a sequential task. In both tasks, only race-specific information needed to be processed in the face stimuli for participants to complete the task successfully and the changes in ethnicity were never associated with changes in identity. Using this paradigm, we found neither an other-race effect for response times nor for response accuracy. In both tasks, the difference in race information between both faces in a pair (the race dissimilarity level) had a significant effect on performance; participants answered more correctly and also faster with increasing dissimilarity level.

In a study by Walker and colleagues [27], participants performed a same/different sequential task using Asian and Caucasian face stimuli and morphs between them. In that experiment, the faces to compare in each trial differed in race and identity. In contrast to our findings, the authors found a clear other-race effect in the accuracy data, participants performed better for same-race faces, while they found no effect in RT data, as we did. The differences between their experiments and ours are that they compared faces differing in race and identity and that they used a different task. In one of our previous studies on the perception of identity and sex in Caucasian faces [21] participants compared faces differing in sex but not in identity in both

a same/different task (as in the Walker & Tanaka study) and a discrimination task (as in our present study) and gave in both tasks qualitatively identical results. This suggests that the use of different tasks cannot explain the qualitative discrepancy between our present findings and those of Walker & Tanaka. We thus argue that the absence or occurrence of an other-race effect – as observed in our study and that of Walker and Tanaka – is mostly due to the absence or presence of an identity change between faces to compare and not because of differences in paradigms. Thus it highlights the importance of processing changing facial identity information alongside race information for the occurrence of the other-race effect.

In another related study, Megreya and his colleagues [28] tested participants in a simultaneous matching task on photographs of people of different ethnicities. They found an other-race effect with their participants showing better matching performance for same-race than other-race faces, suggesting that the other-race effect occurs in the absence of a memory load. In our present study, we found no evidence of an other-race effect in the presence as well as in the absence of a memory load. Together with our study, the findings of Megreya and colleagues [28] and of Walker and Tanaka [27] suggest that an other-race effect can occur, independent of the memory load, as long as an identity change is involved in the faces to compare or to match.

Many models have been proposed for face recognition, one of the most influential ones [29] proposes parallel routes for category processing (e.g. race or sex, called *directed visual processing*), and identity processing. This separation of routes has been questioned in a few studies (for example: [30, 31, 32]). Furthermore, Bruyer and his colleagues [33] have shown that race classification is affected by familiarity with the face stimuli that are judged. Similarly, [14] showed that the emergence of a categorical effect for face sex was linked to participants being familiar with the test identities. In view of these findings and the current study we propose that identity information is not processed separately from other facial information like sex and race.

Acknowledgements We wish to thank Karin Bierig for her help in running this experiment. This study was supported by the Max Planck Society, by the WCU (World Class University) program and by the Brain Korea 21 PLUS Program both funded by the Ministry of Education through the National Research Foundation of Korea. Please address correspondence to I. Bülthoff (isabelle.buelthoff@tuebingen.mpg.de).

References

1. Moses Y, Adini Y, Ullman S (1994) Face recognition: the problem of compensating for changes in illumination direction. Lect Notes Comput Sci 800:286–296
2. Cosmides L, Tooby J, Krurzban R (2003) Perceptions of race. Trends Cogn Sci 7(4):173–179
3. Jenkins R, White D, Van Montfort X, Burton AM (2011) Variability in photos of the same face. Cognition 121(3):313–323
4. Zhou X, Laurence S, Mondloch C (2014) They all look different to me: within-person variability affects identity perception for other-race faces more than own-race faces. J Vis 10(13): 1263, Meeting abstract presented at VSS 2014

5. Hugenberg K, Young SG, Bernstein MJ, Sacco DF (2010) The categorization-individuation model: an integrative account of the other-race recognition deficit. Psychol Rev 117(4):1168–1187
6. Anastasi JS, Rhodes MG (2005) An own-age bias in face recognition for children and older adults. Psychon Bull Rev 12(6):1043–1047
7. Kuefner D, Cassia VM, Picozzi M, Bricolo E (2008) Do all kids look alike ? Evidence for an other-age effect in adults. J Exp Psychol Hum Percept Perform 34(4):811–817
8. Armony JL, Sergerie K (2007) Own-sex effects in emotional memory for faces. Neurosci Lett 426(1):1–5
9. Kret ME, De Gelder B (2012) A review on sex differences in processing emotional signals. Neuropsychologia 50(7):1211–1221
10. Megreya AM, Bindemann M, Havard C (2011) Sex differences in unfamiliar face identification: evidence from matching tasks. Acta Psychol 137(1):83–89
11. Zhao M, Hayward WG, Bülthoff I (2014) Face format at encoding affects the other-race effect in face memory. J Vis 14:1–13
12. Ekman P, Friesen WV (1976) Pictures of facial affect. Consulting Psychologists Press, Palo Alto
13. Bülthoff I, Newell FN (2004) Categorical perception of sex occurs in familiar but not unfamiliar faces. Vis Cogn 11(7):823–855
14. Armann R, Bülthoff I (2012) Male and female faces are only perceived categorically when linked to familiar identities–and when in doubt, he is a male. Vis Res 63:69–80
15. Levin DT (1996) Classifying faces by race: the structure of face categories. J Exp Psychol Learn Mem Cogn 22(6):1364–1382
16. Rhodes G, Ewing L, Hayward WG, Maurer D, Mondloch CJ, Tanaka J (2009) Contact and other-race effects in configural and component processing of faces. Br J Psychol 100(Pt 4):717–728
17. Goldstone RL (1994) Influences of categorization on perceptual discrimination. J Exp Psychol Gen 123(2):178–200
18. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., Los Angeles, pp 187–194
19. Beale JM, Keil FC (1995) Categorical effects in the perception of faces. Cognition 57(3):217–239
20. Levin DT (2000) Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. J Exp Psychol Gen 129(4):559–574
21. Armann R, Bülthoff I (2009) Gaze behavior in face comparison: the roles of sex, task, and symmetry. Atten Percept Psychophys 71(5):1107–1126
22. O'Toole AJ, Deffenbacher KA, Valentin D, Abdi H (1994) Structural aspects of face recognition and the other-race effect. Mem Cogn 22(2):208–224
23. Lee RK, Bülthoff I, Armann R, Wallraven C, Bülthoff HH (2011) The other-race effect is not ubiquitous. J Vis 11(11):626
24. Lee RK, Bülthoff I, Armann R, Wallraven C, Bülthoff HH (2011) Investigating the other-race effect in different face recognition tasks. Iperception 2(4):355–355
25. Hayward WG, Rhodes G, Schwaninger A (2008) An own-race advantage for components as well as configurations in face recognition. Cognition 106(2):1017–1027
26. Meissner CA, Brigham JC (2001) Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol Public Policy Law 7(1):3–35
27. Walker PM, Tanaka J (2003) An encoding advantage for own-race versus other-race faces. Perception 32(9):1117–1125
28. Megreya AM, White D, Burton AM (2011) The other-race effect does not rely on memory: evidence from a matching task. Q J Exp Psychol (Hove) 64(8):1473–1483
29. Bruce V, Young AW (1986) Understanding face recognition. Br J Psychol 77(3):305–327

30. Schweinberger SR, Soukup GR (1998) Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. J Exp Psychol Hum Percept Perform 24(6):1748–1765
31. Baudouin J-Y, Tiberghien G (2002) Gender is a dimension of face recognition. J Exp Psychol Learn Mem Cogn 28(2):362–365
32. Ganel T, Goshen-Gottstein Y (2002) Perceptual integrality of sex and identity of faces: further evidence for the single-route hypothesis. J Exp Psychol Hum Percept Perform 28(4):854–867
33. Bruyer R, Leclere S, Quinet P (2004) Ethnic categorisation of faces is not independent of face identity. Perception 33(2):169–179

Part IV
Early Diagnosis and Prediction of Neural Diseases

Chapter 11
Functional Neuromonitoring in Acquired Head Injury

Hakseung Kim, Young-Tak Kim, and Dong-Joo Kim

Abstract Patients with acquired head injury require accurate and rapid diagnosis regarding their neurophysiological status. As a timely detection of the neuropathological changes in injured brain is important for patient management, a real-time neurological monitoring is common procedure performed in a neurointensive care unit. The neuromonitoring is conducted via acquisition and analysis of various physiological parameters, such as intracranial pressure, cerebral perfusion pressure, intracranial compliance, cerebral autoregulatory capacity, cerebral oxygenation, etc. This article introduces major concepts and parameters in describing the neurological condition of head injured patients. Engineers and scientists who are interested in inter-disciplinary research with neuro-intensivists or neurosurgeons are the intended audience of this article.

Keywords Acquired head injury • Intracranial pressure • Neurointensive care • Neuromonitoring • Traumatic brain injury

Abbreviations

ABP	arterial blood pressure
AHI	acquired head injury
AMP	pulse amplitude of ICP
CA	cerebral auto-regulation
CBF	cerebral blood flow
Cox	correlation coefficient between oxygen saturation and CPP
CPP	cerebral perfusion pressure
CSF	cerebrospinal fluid
CT	computed tomography
CVR	cerebrovascular resistance

H. Kim • Y.-T. Kim • D.-J. Kim (✉)
Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, South Korea
e-mail: dongjookim@korea.ac.kr

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_11

HVx	correlation between rTHb and mean arterial blood pressure
ICP	intracranial pressure
LLA	lower limits of auto-regulation
MRA	magnetic resonance angiography
MRI	magnetic resonance imaging
Mx	index of auto-regulation changes in CPP and the velocity of CBF
NIRS	near infrared spectroscopy
Optimal CPP	optimal cerebral perfusion pressure
PET	positron emission tomography
PRx	pressure reactivity index
PVC	pressure-volume curve
RAP	pressure-volume compensatory index
rTHb	relative total hemoglobin
TBI	traumatic brain injury
TCD	transcranial Doppler
ULA	upper limits of auto-regulation

11.1 Introduction

Acquired head injury (AHI) refers to any non-congenital head injury, with traumatic brain injury (TBI) and stroke as the major causes of AHI (hereafter, 'AHI' refers to TBI and stroke). TBI and stroke are both leading causes of death worldwide [1, 2]. With a few exceptions, AHI patients need prompt care and treatment in an intensive care unit (preferably in a specialized neurointensive care unit) [3].

In a severe AHI, the injury to the patient is biphasic and includes the primary injury that is directly responsible for the initial neurological damage, and the secondary injury that result from a cascade of symptoms caused by the initial damage. Although the primary injury itself cannot be prevented, the extent of the damage due to the secondary injury can be minimized with rapid assessment and proper treatment. In fact, recent studies suggest that neuronal cell death in AHIs, which has long been considered to be the result of the primary injury and, therefore, difficult to treat, actually happens within a few hours after the incident [4]. Thus, in an AHI, a rapid assessment of the neurological condition is the key to obtaining a favorable outcome.

Despite recent advancements, pharmacological interventions for AHIs, especially in TBIs, are effective during the sub-acute or chronic phases [5], but not in the acute phase. Currently, management of the acute phase of AHIs (i.e., during intensive care) focuses on minimizing the secondary injuries [6, 7]. This can be best achieved by real-time neurological monitoring. In addition to the common monitoring of physiological parameters in intensive care, i.e., heart rate and respiratory monitoring, further information is required for the proper treatment of AHI patients in neurointensive care. Important parameters include intracranial pressure (ICP), cerebral perfusion pressure (CPP), intracranial compliance, degree of cerebral auto-regulation, and cerebral oxygenation. This article aims to provide lay descriptions of important concepts and parameters in neurointensive care, as

well as how to measure them, for computer scientists, engineers, or neuroscientists who are interested in developing software or hardware tools to support clinical decision-making in neurointensive care.

11.2 Intracranial Pressure and the Monro-Kellie Doctrine

In the acute phase of AHI, ICP is often elevated. ICP literally refers to the pressure inside the skull (cranium), and is considered to be one of the most important parameters in neurointensive care, as the abnormal increase of ICP indicates a deterioration of neurological status. In clinical practice, ICP is measured at various places, such as the lateral ventricles or subarachnoid spaces (Fig. 11.1). The 'normal' value of ICP is considered to be below 15 mmHg in adults, although slight changes may occur due to changes in body position (Table 11.1).

Elevated ICP is dangerous because the brain is enclosed by a rigid cranium. The Monro-Kellie doctrine (named after Dr. Kellie who proposed the doctrine, and his

Fig. 11.1 Invasive monitoring of intracranial pressure. *Iv* intraventricular, *Ip* intraparenchymal, *Sa* subarachnoid space, *Ed* epidural and *Sd* subdural [8]

Table 11.1 Normal intracranial pressure (ICP) value, measured at the foramina of Monro, supine position [9–14]

Age group	Normal ICP values
Infant	<7.5 mmHg
Child	< 10 mmHg
Adult	< 15 mmHg

teacher) states the intracranial system is enclosed in a non-expandable cranium and the brain parenchyma is nearly incompressible. Indeed, except for a short period early in life, the cranium is non-expandable, and the brain parenchyma is nearly incompressible because it consists mostly of water. This means that the three major entities of the intracranial space, e.g., the cerebrovascular bed, cerebrospinal fluid (CSF), and the brain, must share a single room. Considering the fact that the volume inside the cranium does not change, the sum of the volume of these three entities must be constant, if the pressure inside the cranium (ICP) is to remain constant. In other words, elevated ICP reflects the abnormal increase in volume of at least one intracranial entity.

11.3 Intracranial Pressure in Acquired Head Injury

Immediately after a primary insult to the brain, a disturbance in brain homeostasis will occur [15]. These changes will bring several pathologic events, but the most important one is cell swelling (cerebral swelling/neuronal swelling). As per the Monro-Kellie doctrine, the volume expansion of one component will result in decreases in the volumes of the other two components. This doctrine precisely describes what happens in the intracranial space after parenchymal cells begin to swell; swelling of the parenchymal component will increase ICP, and increase in ICP will decrease the cerebral perfusion pressure (CPP; defined as the mean arterial blood pressure minus ICP). A decrease in CPP decreases cerebral blood flow (CBF); thus, a decrease in CPP effectively results in decreases in the supply of oxygen and glucose that are needed for basic cellular metabolism. In the absence of a proper intervention, these series of events will cause cerebral ischemia. While ischemic insult poses a serious threat to any living tissue, an ischemic insult to cerebral tissue is particularly dangerous for obvious reasons. Because of its devastating effect, the prevention of an ischemic insult is considered to be the primary goal in neurointensive care [16]. However, an abnormally elevated CPP can induce intracranial hypertension or hyperemia (Fig. 11.2). Control of CPP can be achieved by directly intervening blood pressure or ICP.

11.4 Intracranial Pressure and Intracranial Compliance

The extent of the damage from elevated ICP depends on several factors; however, perhaps the most important factor is the intracranial compliance (defined as the change in volume divided by the change in pressure, or $\triangle V/\triangle P$). If the intracranial compliance is high, a small change in volume will not result in a significant change in pressure, and vice versa. This concept is important because a less compliant brain can be susceptible to slight changes in volume, and, by definition, intracranial compliance can be lowered by increasing ICP [17].

The lowering of intracranial compliance does not necessarily involve an increase in ICP, as other causes, such as decreased tissue or vascular compliance, can

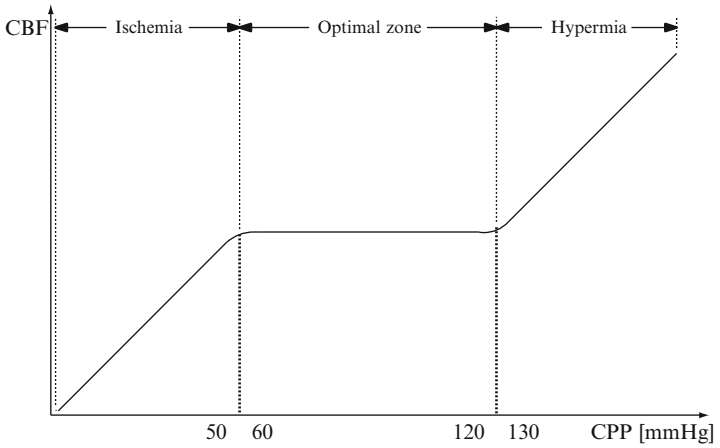

Fig. 11.2 Three different cerebral perfusion pressure (CPP) zones related to clinical status

also cause decreased intracranial compliance. Assessing the change in intracranial compliance can be beneficial, especially in neuro-critical care.

11.5 Cerebral Autoregulation

As stated thus far, an increase in ICP results in significant neurological deterioration, mainly due to decreases in CPP and CBF. To prevent this from happening, our body has the capacity to maintain a constant CBF in the face of moderate changes in CPP. This function is called cerebral auto-regulation (CA). The CBF can be regulated by actively adjusting arteriole diameter, which causes cerebrovascular resistance (CVR) [18, 19]. Nonetheless, the capacity of CA is not limitless; there are lower and upper limits of auto-regulation (LLA and ULA, respectively) [20]. Under normal physiological conditions, CA can cope with a significant degree of fluctuation in the CPP due to the change in ICP. However, patients with AHI often suffer from a failure of CA. The failure of CA can further aggravate a patient's neurological deterioration.

11.6 Beyond Intracranial Pressure Monitoring

Despite the fact that ICP monitoring is associated with favorable outcomes in AHI patients [21–24], the prognostic value of ICP monitoring is still controversial [25, 26]. Indeed, an elevated ICP does not always lead to worsening of the patient; although an abnormally high ICP can be helpful to anticipate mortality, it seems that the ICP value itself cannot provide any additional insights [27, 28]. With advancements in multimodal neuromonitoring, several important parameters have been devised to enhance the diagnostic and prognostic value of ICP.

11.6.1 Pulse Amplitude of ICP (AMP)

As with other waveforms, ICP can be converted to the frequency domain using a Fourier transform. In the frequency spectrum of ICP, the harmonic frequency, in accordance with the frequency of the heartbeat, is called the 'fundamental frequency in ICP'. The amplitude of this component is defined as the pulse amplitude of ICP (AMP). AMP is positively correlated with mean ICP. For instance, an elevation of ICP induces an increase in intracranial capacity as a result of the stroke volume of the heart. Consequently, this increases the AMP (Fig. 11.3). AMP is an important parameter itself. It is used as a supplementary concept, together with other compensatory parameters derived from the ICP waveform [25].

11.6.2 Pressure-Volume Compensatory Index (RAP)

The pressure-volume compensatory index (RAP) is obtained by simultaneously calculating the correlation coefficient between the mean ICP and AMP [29] (Fig. 11.3). RAP is relevant to the pressure-volume curve (PVC) [30, 31], which was derived by the Monro-Kellie doctrine. RAP can be applied to check the condition of compensatory reserve. The RAP value ranges from −1 to 1. A RAP close to 0 indicates a good compensatory mechanism. If the RAP is close to 1, the ICP changes greatly despite a small increase in volume [25–29]. TBI patients often suffer from cerebral edema as a result of head injury. In the case of TBI, the RAP is usually close to 1. RAP can be utilized for various purposes. For example, the effect of a decompressive craniectomy can be confirmed by RAP [32].

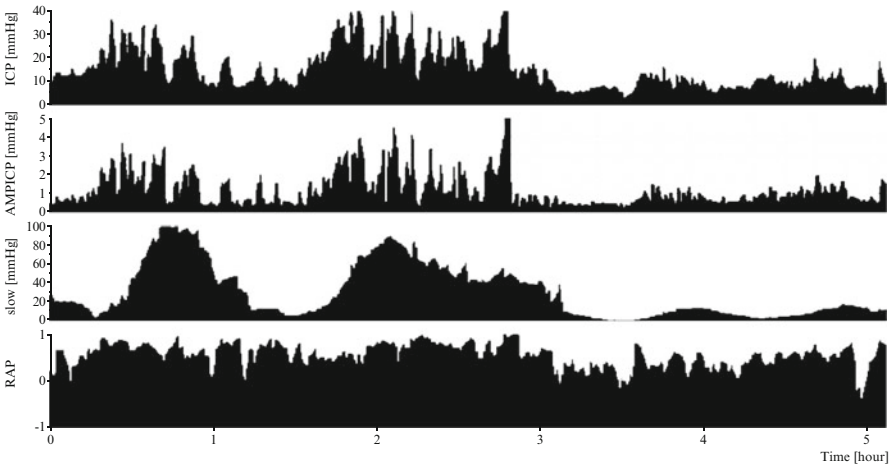

Fig. 11.3 Example of overnight monitoring of intracranial pressure (ICP). Notice the relationship between the raw ICP, the pulse amplitude of ICP (AMP), and the pressure-volume compensation index (RAP)

11.6.3 Pressure Reactivity Index (PRx)

CA is estimated by assessing ICP response in accordance with continuous changes in the ICP [33]. In normal physiologic condition, CBF is maintained in the normal range. However, a deteriorated CA responds slowly to the change in ABP in TBI patients. Consequently, this difference is quantified by the pressure reactivity index (PRx), which is calculated from the correlation between the continuous change in ABP and ICP (Fig. 11.4). A PRx below 0 indicates normal CA, while a PRx above 0 indicates deteriorated CA. According to recent research based on adults, a PRx above 0.3 is strongly correlated with mortality [34], and an equivalent result was shown in a similar study ($n = 193$, $p < 0.0001$) [35]. A study based on pediatric subjects showed that the difference in mean PRx between surviving and deceased subjects is about 0.61 [36].

11.6.4 Optimal Cerebral Perfusion Pressure (Optimal CPP)

Recent studies examined the relationship between PRx and optimal CPP [36–38]. Optimal CPP is a CPP that is in the normal range, and which does not result in ischemia or congestion of the brain (Fig. 11.4). It can be obtained after several hours of monitoring. A previous study of optimal CPP in 144 subjects used the relationship between PRx and CPP to establish that the optimal CPP is about 60 % [39].

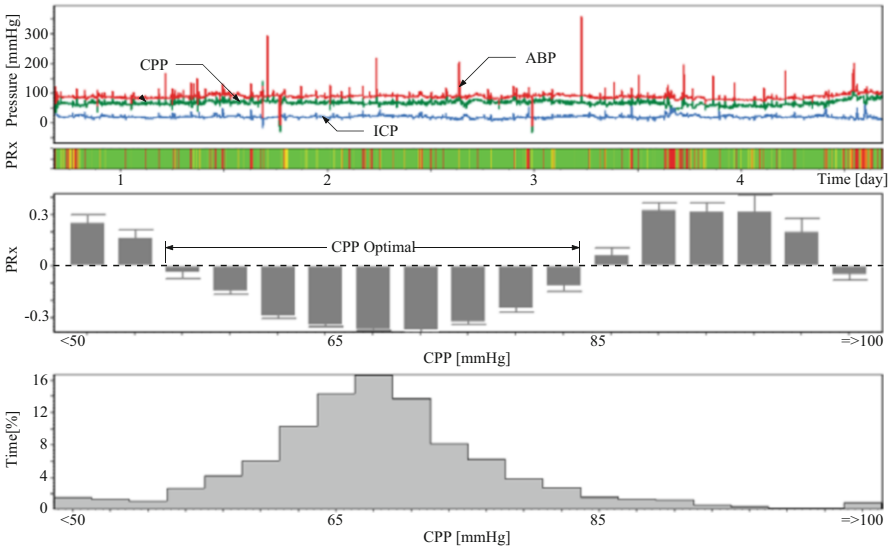

Fig. 11.4 Optimal CPP zone, formulated by the relationship between the pressure reactivity index (PRx) and cerebral perfusion pressure (CPP)

11.7 Monitoring Cerebral Auto-Regulation

11.7.1 Transcranial Doppler

CBF was first measured in 1945, and has remained the main target for monitoring. There are various procedures that measure CBF, primarily perfusion computed tomography (CT), magnetic resonance angiography (MRA), positron emission tomography (PET), oxygenation, transcranial Doppler (TCD), laser Doppler, and jugular vein oximetry. Especially, TCD and jugular vein oximetry are widely used in intensive care units. TCD estimates CBF using the Doppler effect caused by the movement of red blood cells [7]. It is widely known that the measurement of CBF using TCD is an estimate value [40]. TCD is a length procedure, and various factors influence the estimated CBF value [41]. Researchers should keep in mind that it is necessary to eliminate signal artifacts when analyzing TCD. However, there is no method that automatically eliminates artifacts and, thus, these procedures are labor-intensive. TCD is noninvasive, simple, low cost, and painless for patients. Above all, TCD can serve multiple purposes [42–44]. Consequently, methods for measuring ICP or CBF using TCD have been investigated.

Mx is an index of auto-regulation. It is derived from the correlation coefficient between the voluntary change in CPP and the velocity of CBF [45]. Mx has the same purpose as PRx. In the same manner, Mx below 0 indicates normal CA, and Mx above 0 indicates a loss of CA.

A previous study, which compared PRx and Mx in the same population, reported a strong relationship between them ($R^2 = 0.36$, $p < 0.001$). However, it is also reported that Mx is harder to measure than PRx. Despite this shortcoming, Mx has remained an index that complements the weakness of PRx via its noninvasiveness [45].

11.7.2 Near-Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) is noninvasive method that measures oxygen saturation in a local area of the brain. The area to be measure is illuminated by near-infrared rays from a light sensor on the scalp. Hemoglobin and myoglobin have different absorption spectra that are dependent on the degree of oxidation-reduction. The scalp and skull have low absorption rates of near-infrared rays (700–1000 nm); thus, the rays from NIRS reach a depth of several centimeters through the scalp [46]. Although NIRS measures a localized area and is subject to noise caused by hemoglobin in the scalp and hematoma, it has an excellent spatio-temporal resolution and is noninvasiveness.

Measurement of relative total hemoglobin (rTHb) using NIRS is a universal method to assess oxygen saturation. However, it is insufficient to monitor changes in various lesions in TBI patients. To overcome this limitation, Cox and HVx were

proposed. Cox is defined as the correlation coefficient between oxygen saturation and CPP. Similarly, HVx is derived as the correlation between rTHb and mean arterial blood pressure.

Cox and HVx are also used to assess CA, as are PRx and Mx. Conditions in which the Cox and HVx are close to 0 or negative indicates a normal CA. If CA is deteriorated by a disorder (e.g., hypotension), these parameters show positive values. Many researchers have attempted to utilize these parameters, but its accuracy is still controversial [47–49].

11.8 Monitoring Intracranial Compliance

The importance of monitoring intracranial compliance has been well accepted, and there are various methods by which it can be measured. Previously, the monitoring of intracranial compliance was done manually, by injecting and then withdrawing a bolus of mock CSF [50–53]. Although such methods are still used, current ICP monitoring devices are often equipped with an automated compliance monitoring capacity (Fig. 11.5).

Other than the use of specialized devices, the raw recordings of ICP can also enable the assessment of intracranial compliance, and this can be achieved via the morphological analysis of the ICP waveforms of. The ICP waveform has three distinctive peaks, named P1, P2, and P3 in the order of appearance (Fig. 11.6) [55].

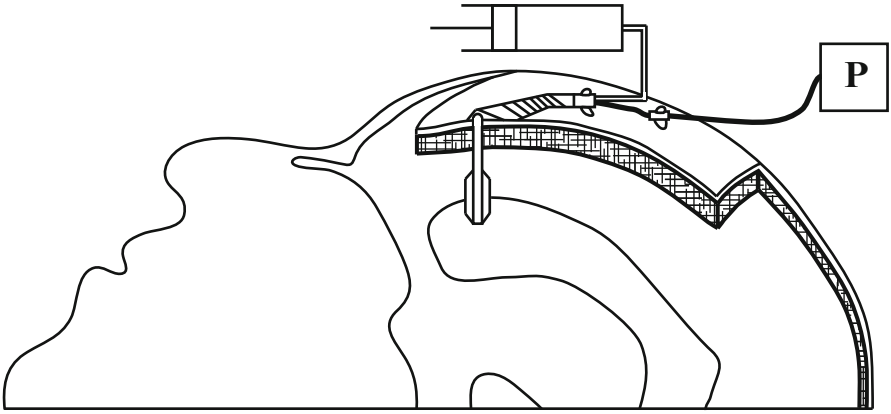

Fig. 11.5 Diagrammatic representation of the Spiegelberg Compliance monitoring device. The system incorporates a piston-based pump for accurately adding and removing small volumes of air (0.2 ml) into a double lumen catheter placed within a ventricle. A strain-gauge pressure transducer (P) connects to the fluid-controlled lumen of the catheter to measure the small perturbations in intracranial pressure (ICP) resulting from the addition and removal of small volumes to the balloon. The figure was adapted and modified from the original work of Piper et al. [54]

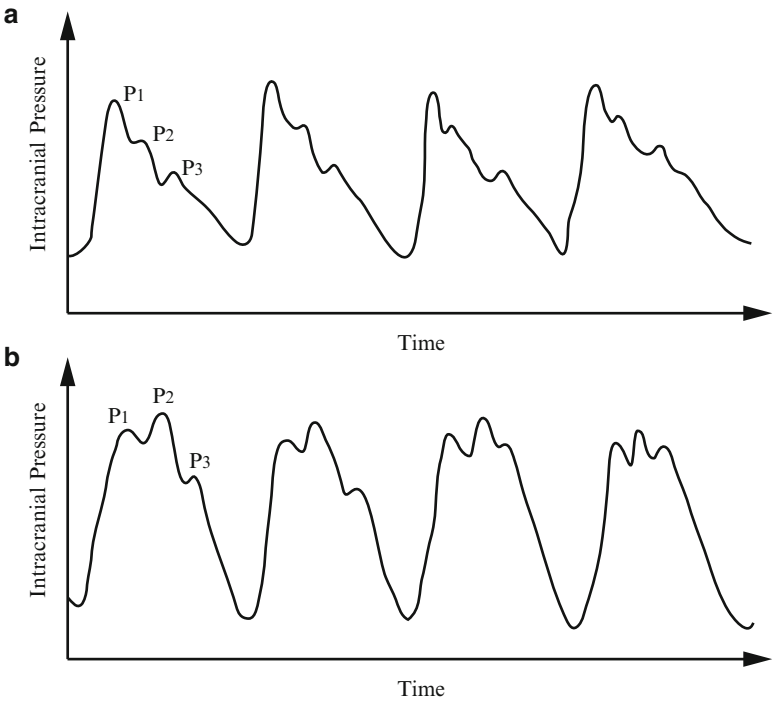

Fig. 11.6 Typical intracranial pressure (ICP) waveforms in an intracranial system with (**a**) high compliance and (**b**) decreased compliance

The waveform of elevated ICP typically shows a decreased P1 amplitude in P1 and an increased P2 amplitude [56], and this phenomenon has long been considered as indicative of decreased intracranial compliance [55, 57–59].

The use of invasive devices can provide accurate and continuous monitoring of intracranial compliance; however, it also poses a non-negligible risk of infection or hemorrhage. Recently, non-invasive methods involving magnetic resonance imaging (MRI) have been proposed [60, 61]. However, these methods share some significant limitations, namely that MRI only allows a temporary assessment for a very short duration. Because of the inherent nature of MRI, it cannot provide a continuous, real-time assessment of intracranial compliance.

11.9 Conclusion

Neurointensive care can help prevent the further aggravation of injury due to the worsening of secondary injuries in AHI patients, and ICP monitoring plays an essential role in this process. Nonetheless, the absolute value of ICP itself may

not be useful, other than to anticipate mortality. This limitation can be overcome by multimodal neuromonitoring. Multimodal neuromonitoring not only enables a multilateral assessment of a patient's neurological condition, but also enhances the diagnostic and prognostic value of ICP. The remaining drawback of monitoring ICP lies in its invasiveness; further research should focus on developing reliable, highly accurate, noninvasive ICP measurement techniques.

Acknowledgement The authors acknowledge the use of text from the prior dissertation "Clinical and engineering models of brain compliance and deformation associated with neurological disorders."

References

1. Towfighi A, Saver JL (2011) Stroke declines from third to fourth leading cause of death in the United States: historical perspective and challenges ahead. Stroke 42(8):2351–2355. doi:10.1161/strokeaha.111.621904 [published Online First: Epub Date]
2. Gean AD, Fischbein NJ (2010) Head trauma. Neuroimaging Clin N Am 20(4):527–556. doi:10.1016/j.nic.2010.08.001 [published Online First: Epub Date]
3. Jones PA, Andrews PJ, Midgley S et al (1994) Measuring the burden of secondary insults in head-injured patients during intensive care. J Neurosurg Anesthesiol 6(1):4–14
4. Reilly PL (2001) Brain injury: the pathophysiology of the first hours. 'Talk and Die revisited'. J Clin Neurosci Off J Neurosurg Soc Australas 8(5):398–403. doi:10.1054/jocn.2001.0916 [published Online First: Epub Date]
5. Talsky A, Pacione LR, Shaw T et al (2010) Pharmacological interventions for traumatic brain injury. B C Med J 53(1): 26–31
6. Mendelow AD, Crawford PJ (2005) Primary and secondary brain injury. In: Bullock PLRR (ed) Head injury – pathophysiology and management, 2nd edn. Hodder Education, Great Britain
7. Tisdall MM, Smith M (2007) Multimodal monitoring in traumatic brain injury: current status and future directions. Br J Anaesth 99(1):61–67. doi:10.1093/bja/aem143 [published Online First: Epub Date]
8. Weerakkody RA, Czosnyka M, Trivedi RA et al (2009) Intracranial pressure monitoring in head injury. Cambridge University Press, Cambridge/New York/Melbourne/Madrid/Cape Town/Singapore/Sao Paulo/Delhi
9. Gilland O, Tourtellotte WW, O'Tauma L et al (1974) Normal cerebrospinal fluid pressure. J Neurosurg 40(5):587–593. doi:10.3171/jns.1974.40.5.0587 [published Online First: Epub Date]
10. Langfitt TW, Weinstein JD, Kassell NF et al (1964) Transmission of increased intracranial pressure: II. Within the supratentorial space. J Neurosurg 21:998–1005. doi:10.3171/jns.1964.21.11.0998 [published Online First: Epub Date]
11. Lundberg N (1960) Continuous recording and control of ventricular fluid pressure in neurosurgical practice. Acta Psychiatr Scand Suppl 36(149):1–193
12. Magnaes B (1976) Body position and cerebrospinal fluid pressure: Part 2: clinical studies on orthostatic pressure and the hydrostatic indifferent point. J Neurosurg 44(6): 698–705. doi:10.3171/jns.1976.44.6.0687 [published Online First: Epub Date]
13. Magnaes B (1976) Body position and cerebrospinal fluid pressure: Part 1: Clinical studies on the effect of rapid postural changes. J Neurosurg 44(6):687–697. doi:10.3171/jns.1976.44.6.0687 [published Online First: Epub Date]
14. Welch K (1980) The intracranial pressure in infants. J Neurosurg 52(5):693–699. doi:10.3171/jns.1980.52.5.0693 [published Online First: Epub Date]

15. McAuley D (2009) Early phase care of patients with moderate and severe head injury. In: Whitfield PC, Thomas EO, Summers F, Whyte M, Hutchinson PJ (eds) Head injury: a multidisciplinary approach. Cambridge University Press, Cambridge, pp 62–78
16. Smith M (2009) Principle of head injury intensive care management. In: Whitfield PC, Thomas EO, Summers F, Whyte M, Hutchinson PJ (eds) Head injury: a multidisciplinary approach. Cambridge University Press, Cambridge, pp 79–86
17. Wagshul ME, Eide PK, Madsen JR (2011) The pulsating brain: a review of experimental and clinical studies of intracranial pulsatility. Fluids Barriers CNS 8(1):5. doi:10.1186/2045-8118-8-5 [published Online First: Epub Date]
18. Fog M (1938) The relationship between the blood pressure and the tonic regulation of the pial arteries. J Neurol Psychiatry 1(3):187–197
19. Symon L, Held K, Dorsch NW (1973) A study of regional autoregulation in the cerebral circulation to increased perfusion pressure in normocapnia and hypercapnia. Stroke 4(2):139–147
20. Budohoski KP, Czosnyka M, Kirkpatrick PJ et al (2013) Clinical relevance of cerebral autoregulation following subarachnoid haemorrhage. Nat Rev Neurol. doi:10.1038/nrneurol.2013.11 [published Online First: Epub Date]
21. Vincent JL, Berre J (2005) Primer on medical management of severe brain injury. Crit Care Med 33(6):1392–1399
22. Dutton RP, McCunn M (2003) Traumatic brain injury. Curr Opin Crit Care 9(6):503–509
23. Helmy A, Vizcaychipi M, Gupta AK (2007) Traumatic brain injury: intensive care management. Br J Anaesth 99(1):32–42. doi:10.1093/bja/aem139 [published Online First: Epub Date]
24. Patel HC, Bouamra O, Woodford M et al (2005) Trauma audit and research network: trends in head injury outcome from 1989 to 2003 and the effect of neurosurgical care: an observational study. Lancet 366:1538–1544
25. Czosnyka M, Pickard JD (2004) Monitoring and interpretation of intracranial pressure. J Neurol Neurosurg Psychiatry 75(6):813–821
26. Cremer OL, van Dijk GW, van Wensen E et al (2005) Effect of intracranial pressure monitoring and targeted intensive care on functional outcome after severe head injury. Crit Care Med 33(10):2207–2213
27. Resnick DK, Marion DW, Carlier P (1997) Outcome analysis of patients with severe head injuries and prolonged intracranial hypertension. J Trauma 42(6):1108–1111
28. Balestreri M, Czosnyka M, Hutchinson P et al (2006) Impact of intracranial pressure and cerebral perfusion pressure on severe disability and mortality after head injury. Neurocrit Care 4(1):8–13. doi:10.1385/NCC:4:1:008 [published Online First: Epub Date]
29. Kim DJ, Czosnyka Z, Keong N et al (2009) Index of cerebrospinal compensatory reserve in hydrocephalus. Neurosurgery 64(3):494–501; discussion 01–2 doi:10.1227/01.NEU.0000338434.59141.89 [published Online First: Epub Date]
30. Lofgren J, von Essen C, Zwetnow NN (1973) The pressure-volume curve of the cerebrospinal fluid space in dogs. Acta Neurol Scand 49(5):557–574
31. Avezaat CJ, van Eijndhoven JH, Wyper DJ (1979) Cerebrospinal fluid pulse pressure and intracranial volume-pressure relationships. J Neurol Neurosurg Psychiatry 42(8):687–700
32. Whitfield PC, Patel H, Hutchinson PJ et al (2001) Bifrontal decompressive craniectomy in the management of posttraumatic intracranial hypertension. Br J Neurosurg 15(6):500–507
33. Balestreri M, Czosnyka M, Steiner LA et al (2004) Intracranial hypertension: what additional information can be derived from ICP waveform after head injury? Acta Neurochir 146(2):131–141. doi:10.1007/s00701-003-0187-y [published Online First: Epub Date]
34. Timofeev I, Czosnyka M, Nortje J et al (2008) Effect of decompressive craniectomy on intracranial pressure and cerebrospinal compensation following traumatic brain injury. J Neurosurg 108(1):66–73. doi:10.3171/JNS/2008/108/01/0066 [published Online First: Epub Date]
35. Balestreri M, Czosnyka M, Steiner LA et al (2005) Association between outcome, cerebral pressure reactivity and slow ICP waves following head injury. Acta Neurochir Suppl 95:25–28

36. Brady KM, Shaffner DH, Lee JK et al (2009) Continuous monitoring of cerebrovascular pressure reactivity after traumatic brain injury in children. Pediatrics 124(6):e1205–e1212. doi:10.1542/peds.2009-0550 [published Online First: Epub Date]
37. Zweifel C, Lavinio A, Steiner LA et al (2008) Continuous monitoring of cerebrovascular pressure reactivity in patients with head injury. Neurosurg Focus 25(4):E2. doi:10.3171/FOC.2008.25.10.E2 [published Online First: Epub Date]
38. Jaeger M, Dengl M, Meixensberger J et al (2010) Effects of cerebrovascular pressure reactivity-guided optimization of cerebral perfusion pressure on brain tissue oxygenation after traumatic brain injury. Crit Care Med 38(5):1343–1347. doi:10.1097/CCM.0b013e3181d45530 [published Online First: Epub Date]
39. Steiner LA, Czosnyka M, Piechnik SK et al (2002) Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury. Crit Care Med 30(4):733–738
40. Timofeev I, Helmy A, Silva EJ et al (2009) Multimodality monitoring in head injury. In: Whitfield PC, Thomas EO, Summers F, Whyte M, Hutchinson PJ (eds) Head injury: a multidisciplinary approach. Cambridge University Press, Cambridge
41. Czosnyka M, Smielewski P, Lavinio A et al (2008) An assessment of dynamic autoregulation from spontaneous fluctuations of cerebral blood flow velocity: a comparison of two models, index of autoregulation and mean flow index. Anesth Analg 106(1):234–239, table of contents doi:10.1213/01.ane.0000295802.89962.13 [published Online First: Epub Date]
42. Czosnyka M, Matta BF, Smielewski P et al (1998) Cerebral perfusion pressure in head-injured patients: a noninvasive assessment using transcranial Doppler ultrasonography. J Neurosurg 88(5):802–808. doi:10.3171/jns.1998.88.5.0802 [published Online First: Epub Date]
43. Springborg JB, Frederiksen HJ, Eskesen V et al (2005) Trends in monitoring patients with aneurysmal subarachnoid haemorrhage. Br J Anaesth 94(3):259–270. doi:10.1093/bja/aei004 [published Online First: Epub Date]
44. Bellner J, Romner B, Reinstrup P et al (2004) Transcranial Doppler sonography pulsatility index (PI) reflects intracranial pressure (ICP). Surg Neurol 62(1):45–51; discussion 51 doi:10.1016/j.surneu.2003.12.007 [published Online First: Epub Date]
45. Lavinio A, Schmidt EA, Haubrich C et al (2007) Noninvasive evaluation of dynamic cerebrovascular autoregulation using Finapres plethysmograph and transcranial Doppler. Stroke 38(2):402–404. doi:10.1161/01.STR.0000254551.92209.5c [published Online First: Epub Date]
46. Ercole A, Gupta AK (2011) Cerebral oxygenation. In: Basil F, Matta DKM, Smith M (eds) Core topics in neuroanaesthesia and neurointensive care. Cambridge University Press, Cambridge
47. Brady KM, Lee JK, Kibler KK et al (2008) Continuous measurement of autoregulation by spontaneous fluctuations in cerebral perfusion pressure: comparison of 3 methods. Stroke 39(9):2531–2537. doi:10.1161/STROKEAHA.108.514877 [published Online First: Epub Date]
48. Brady KM, Lee JK, Kibler KK et al (2007) Continuous time-domain analysis of cerebrovascular autoregulation using near-infrared spectroscopy. Stroke 38(10):2818–2825. doi:10.1161/STROKEAHA.107.485706 [published Online First: Epub Date]
49. Lee JK, Kibler KK, Benni PB et al (2009) Cerebrovascular reactivity measured by near-infrared spectroscopy. Stroke 40(5):1820–1866. doi:10.1161/STROKEAHA.108.536094 [published Online First: Epub Date]
50. Miller JD, Garibi J, Pickard JD (1973) Induced changes of cerebrospinal fluid volume: effects during continuous monitoring of ventricular fluid pressure. Arch Neurol 28(4):265–269
51. Ryder HW, Espey FF, Kimbell FD et al (1953) The mechanism of the change in cerebrospinal fluid pressure following an induced change in the volume of the fluid space. J Lab Clin Med 41(3):428–435
52. Shulman K, Marmarou A (1971) Pressure-volume considerations in infantile hydrocephalus. Dev Med Child Neurol 13(s25):90–95

53. Raabe A, Czosnyka M, Piper I et al (1999) Monitoring of intracranial compliance: correction for a change in body position. Acta Neurochir 141(1):31–36
54. Piper I, Spiegelberg A, Whittle I et al (1999) A comparative study of the Spiegelberg compliance device with a manual volume-injection method: a clinical evaluation in patients with hydrocephalus. Br J Neurosurg 13(6):581–586
55. Cardoso ER, Rowan JO, Galbraith S (1983) Analysis of the cerebrospinal fluid pulse wave in intracranial pressure. J Neurosurg 59(5):817–821
56. Novák D, Cuesta-Frau D, Aboy M, et al (2004) Clustering of intracranial pressure using hidden Markov models. In: 17th European meetings on cybernetics and systems research, Vienna, 13–16 April 2004
57. Kirkness CJ, Mitchell PH, Burr RL et al (2000) Intracranial pressure waveform analysis: clinical and research implications. J Neurosci Nurs 32(5):271–277
58. Robertson CS, Narayan RK, Contant CF et al (1989) Clinical experience with a continuous monitor of intracranial compliance. J Neurosurg 71(5):673–680
59. Ross N, Eynon C (2005) Intracranial pressure monitoring. Curr Anaesth Crit Care 16(4):255–261
60. Balédent O, Gondry-Jouet C, Meyer M-E et al (2004) Relationship between cerebrospinal fluid and blood dynamics in healthy volunteers and patients with communicating hydrocephalus. Investig Radiol 39(1):45–55
61. Strik C, Klose U, Erb M et al (2002) Intracranial oscillations of cerebrospinal fluid and blood flows: analysis with magnetic resonance imaging. J Magn Reson Imaging 15(3):251–258

Chapter 12
Diagnostic Optical Imaging Technology and Its Principles

Jae-Ho Han

Abstract Analysis of tissue structures is important for investigating pathological changes and diagnosing neural diseases. Recent advances in ophthalmology regarding the diagnosis of neurosensory retinal diseases have introduced optical coherence tomography (OCT) as a near-infrared imaging modality to provide noninvasive and real-time imaging and sensing with ultrahigh resolution for imaging subsurface cross sections of the human retina. This chapter gives a brief overview and the basic principles of this emerging optical imaging modality.

Keywords Optical imaging • Optical coherence tomography • Diagnosis • Retina • Ophthalmology

12.1 Introduction to Optical Coherence Tomography

Effective therapeutic treatment can significantly benefit from early diagnosis and continued monitoring of pathological tissue alterations using high-resolution imaging methods that are neither harmful nor invasive to healthy tissue. Optical coherence tomography (OCT) is a minimally invasive real-time, high-resolution imaging modality that has seen much recent progress and is a promising diagnostic imaging technique [1, 2]. OCT is based on low-coherence optical interferometry, and utilizes ballistic and near-ballistic photons in a turbid medium. The volumetric tissue is illuminated using a scanning optical beam and the backscattered photons can be measured as a function of the path length in the tissue. Thus, OCT systems can be used to obtain depth-resolved *in situ* and *in vivo* cross-sectional and three-dimensional morphological images of biological tissue microstructures, as well as their subsurface depth profiles [3–6]. Figure 12.1 shows sample OCT images of a normal human retina and some common retinal diseases for comparison.

J.-H. Han (✉)
Department of Brain and Cognitive Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 136-701, Republic of Korea
e-mail: hanjaeho@korea.ac.kr

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_12

Table 12.1 lists a comparison between OCT and other imaging modalities, including X-ray computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound sonograms. OCT is analogous to sonograms, with the exception that OCT measures echo reflections of backscattered light rather than backscattered acoustic signals. Although the imaging depth in highly scattering tissue is typically smaller than 1–2 mm, OCT can achieve micrometer-scale axial resolution using

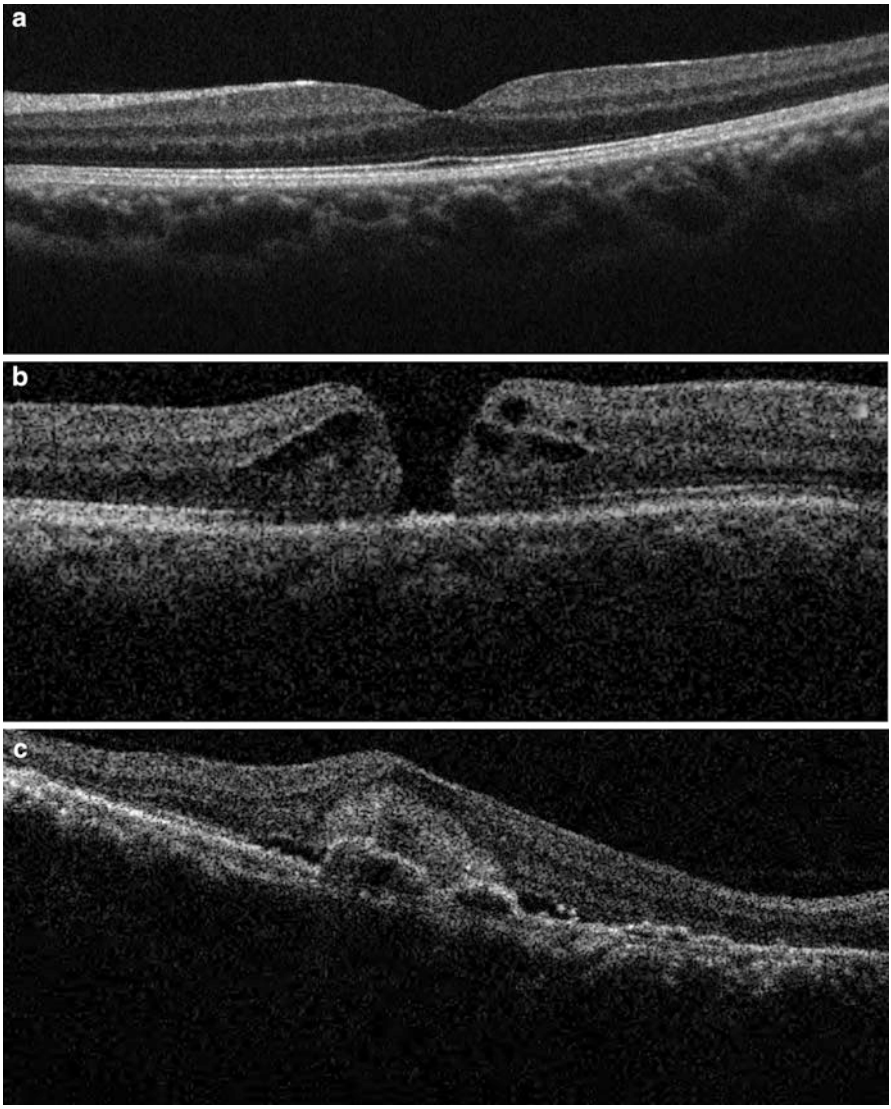

Fig. 12.1 Images obtained using OCT of: (**a**) a normal retina, (**b**) a macular hole, (**c**) age-related macular degeneration, (**d**) an idiopathic epiretinal membrane, and (**e**) a glaucoma patient's retina

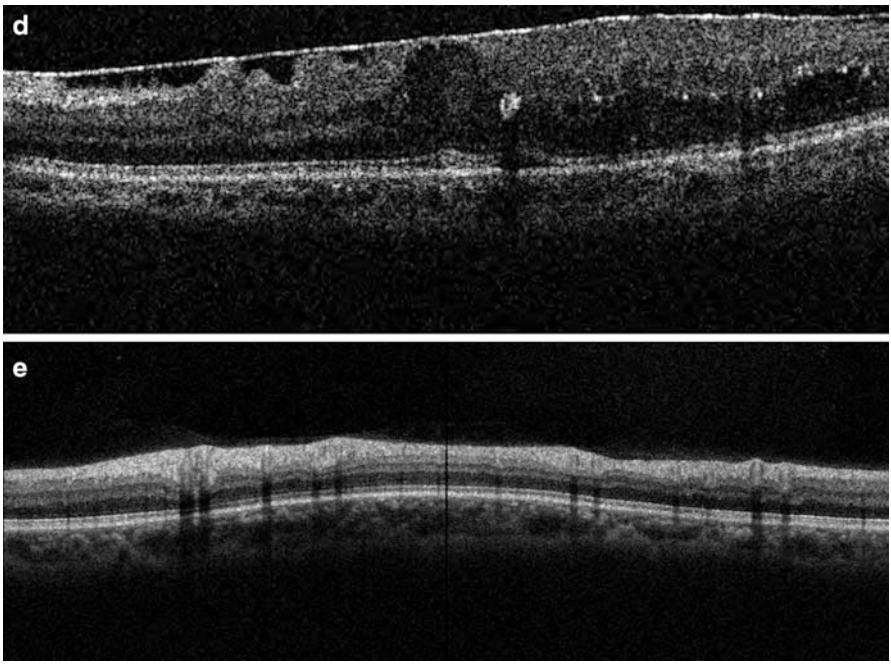

Fig. 12.1 (continued)

Table 12.1 Properties of biomedical imaging modalities

Method	Resolution	Penetration depth	Source of contrast
X-ray Computed Tomography (CT)	2–3 mm	Entire body	Attenuation
Magnetic Resonance Imaging (MRI)	2–3 mm	Entire body	H^+ concentration
Ultrasound Sonogram	500 μm	10–20 cm	Acoustic scattering
Visual Examination	100 μm	Surface	Natural coloring
Optical Coherence Tomography (OCT)	10 μm	2–3 mm	Optical scattering

an ultra-broadband light source. OCT has been widely used in ophthalmology to accurately measure the thickness of the cornea and assess intraocular pathology, and it is rapidly becoming a standard of care used for diagnosing retinal diseases [7, 8]. The rapid adoption of the method in ophthalmology has been partly due to the relative transparency of the relevant ocular tissues for efficiently facilitating the implementation of OCT withnear infrared light. There have also been efforts towards its implementation in areas including cardiology [9], gastroenterology [10], urology [11], dermatology [12], and various delicate microsurgical applications [13].

In the following sections, the basic principles and selected background physics required to understand the implementation and operation of OCT are provided in detail. OCT system configuration for a fiber-optic-based system is succinctly described in Sect. 12.2. The coherence and interference of light are explained in Sect. 12.3 to clarify the concepts associated with light detection mechanisms. Section 12.4 presents the basic steps for analyzing the acquired data for image reconstruction in different domains and Sect. 12.5 presents the interaction of light in at turbid medium to describe the inherent depth-limiting factors of light in the imaging system. A brief summary of this Chapter is provided in Sect. 12.6.

12.2 System Configuration

The schematic configuration of a simple fiber-optic version of an OCT system is shown in Fig. 12.2. A 2×2 fiber-optic coupler is employed to divide and recombine the beams instead of a beam splitter, as in free-space interferometers. This configuration represents the basic structure of dual-arm OCT systems, which involve two optical paths: one for the reference path and one for the sample (object). Here, either the mirror in the reference path is movable along the axial direction or fiber can be stretched to gradually change and precisely match the path difference between the reference and the sample arms.

OCT employs a low-coherence interferometry method by using a light source with a large bandwidth, corresponding to a small coherence length, which requires highly accurate path-length matching between the two arms. Thus, the depth scan is achieved by detecting the positions of the reflected beams in the tissue by scanning the reference arm length. The transverse scan to obtain 2-D and 3-D images can be achieved by moving either the sample or the probe in a lateral direction.

The typical Michelson interferometers used for time-domain (or spatial-domain) analysis have been replaced by the so-called spectral or Fourier-domain OCT. This

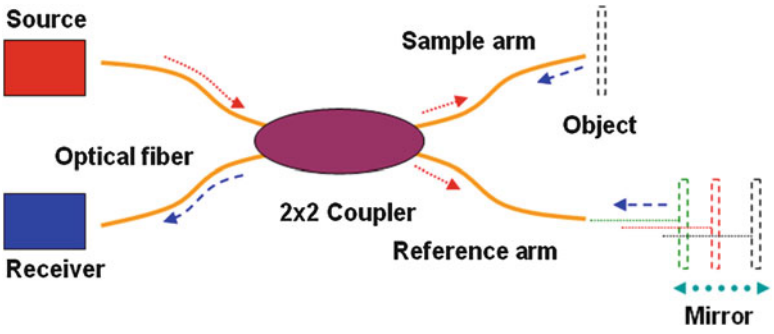

Fig. 12.2 Conventional fiber-optic interferometer-based OCT system. The *arrows* indicate the direction of light

system uses the spectrally dispersed interference signal between the sample and the reference light source to deduce the depth-resolved reflectance by performing a Fourier transform.

12.3 Coherence and Interference

12.3.1 Coherence Length of Light

The light sources used with OCT are based on coherence, which determines the axial (i.e., depth) resolution of the cross-sectional image. Phase changes of non-monochromatic light waves are a stochastic process in which the excited atoms in the light source experience transitions between energy levels and which results in finite harmonic wave trains. The waves statistically maintain their coherence within the average train life (with coherence time τ_0) or, equivalently, are related to the average distance (with coherence length l_c) [14] as

$$l_c = \alpha c \tau_0 = \alpha \frac{c}{\Delta f}, \tag{12.1}$$

where c is the speed of light in free space, Δf is the frequency bandwidth, and α is a constant coefficient representing the statistical distribution of the source (e.g., $\alpha = 0.44$ for a Gaussian source). The conversion between the spectral width at wavelength $\Delta\lambda$ and the bandwidth at frequency Δf is given by

$$\frac{\Delta f}{f} = \frac{\Delta \lambda}{\lambda},\ c = f \cdot \lambda, \tag{12.2}$$

so that the coherence length is related to the wavelength by

$$l_c = \alpha \frac{\lambda^2}{\Delta \lambda}. \tag{12.3}$$

For instance, a perfect monochromatic source ($\Delta\lambda \to 0$) corresponds to an infinite coherence length ($l_c \to \infty$).

In addition, abrupt phase changes reflect the spectral width and the shape of the source. In this case, the interaction of low-coherence waves generates interference fringe patterns only if the path difference $\Delta L = |L_1 - L_2|$ is smaller than the coherence length. The total intensity can be written as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\delta) \left| G(\Delta L) \right|, \tag{12.4}$$

where δ is the phase difference due to the difference in path lengths, which can be written as $\delta = k\Delta L$. $G(\Delta L)$ describes the envelope function of the low-coherence source, and I_1 and I_2 are the intensities of the two interfering waves. The Fourier

transform of this temporal (or spatial) auto-correlation corresponds to the power spectral density. Therefore, the fringe shows a peak if the paths are matched, tapers in either direction if the path difference increases to $\Delta L = \pm l_c$, and vanishes for larger path differences.

12.3.2 Optical Interferometer

An interferometer provides a means to explore the interference of light by studying the fringe patterns that result from the interference between two beams. The interferometer splits an initial beam (amplitude division) into two smaller ones that travel on different optical paths and recombine to produce an interference pattern. The division of the light beam can be performed using a beam splitter so that one beam is transmitted and the other is reflected with equal amplitudes. Both beams are directed to mirrors, are reversed, recombined, and subsequently detected using a single photo detector. When using a beam splitter, at least one of the mirrors should be movable along the optical axis to precisely determine the distance between the two paths within the coherence length of the light source. In the OCT method, the moving mirror acts as a reference arm, and the fixed mirror is the sample arm, where the specimen is placed. Following the distance scan, during which the average power is detected for every position, depth-resolved information can be reconstructed (i.e., an A-mode scan).

12.4 Extraction of Depth Information

12.4.1 Principle of Time-Domain Analysis

In interferometry analysis, the extraction of the path differences is performed based on space or time, so that the mechanical adjustment of the reference path length is indispensable; this process makes the entire system slower. If a spectrometer is used as a receiver instead of a photo detector, the system can be configured in a simpler manner without the mechanical speed limitations. The distance information can be extracted from a Fourier transform of the interference signal using the data obtained with the spectrometer. Therefore, only a lateral scan is required to obtain a two-dimensional cross-sectional image. This method is based on the Wiener-Khintchine theorem, which shows that the power spectrum and its auto-correlation are a Fourier-transform pair that can be expressed as

$$S_{EE}(f) = \int_{-\infty}^{\infty} G_{EE}(\tau) \cdot e^{-i2\pi f\tau} d\tau \leftrightarrow G_{EE}(\tau) = \int_{-\infty}^{\infty} E(t)E^*(t-\tau)\, dt, \quad (12.5)$$

where $G(\tau)$ is the normalized complex coherence function related to the power spectral density of the source according to the Weiner-Khintchine theorem [15], i.e.,

$$G(\tau) = e^{\left[-\left(\frac{\pi \Delta \upsilon \tau}{2\sqrt{\ln 2}}\right)^2\right]} e^{(-i2\upsilon_0 \pi)}, \tag{12.6}$$

where υ_0 is the central frequency of the broadband source and $\Delta\upsilon$ is the half-power bandwidth, which represents the spectral width of the source. For instance, using Fourier-domain analysis, the distance information can be extracted from the interference term in the spectral domain by

$$\begin{aligned} I_{total} &= \int_{-\infty}^{\infty} I(k)dk = \int_{-\infty}^{\infty} \sqrt{I_r(k)I_s(k)} \cos(2kd_i)\, dk \leftrightarrow I(x) = F^{-1}[I_{total}] \\ &= G(x) * \delta(x - 2d_i) = G(x - 2d_i). \end{aligned} \tag{12.7}$$

where F^{-1} is the inverse Fourier transform and d_i represents the position or distance to the ith layer.

12.4.2 Concept of Fourier Domain Conversion

As mentioned above, Fourier analysis can simplify the extraction of depth information with directly using the spectral information instead of using the time domain scanned data. To convert spectral information to an image, the spectral data acquired with the spectrometer are converted from the wavelength space to the wave-vector space (i.e., k-space) using a fast Fourier transform (FFT) to reconstruct the longitudinal profile of the samples. Here, the pixel size of each bin, as obtained from the sampling theorem, is given by

$$\Delta l = \frac{z_s}{N} = \frac{\pi}{\Delta k} = \frac{\lambda_{\max}\lambda_{\min}}{2(\lambda_{\max} - \lambda_{\min})}, \tag{12.8}$$

where N is the total number of elements in the spectrometer detector array, Δk is the spectral width in k-space, and z_s is the imaging depth. For example, for a high-resolution fiber-optic spectrometer that covers the spectral range 700–900 nm and uses a charge-coupled device (CCD) array with 3648 pixels, the equidistant bin corresponds to approximately 1.6 μm. This translates to a maximum imaging depth of approximately 3.5 mm and a spectral resolution of 0.09 nm.

12.5 Light-Matter Interaction and Refractive Index

12.5.1 Characteristics of Light Propagation in Matter

Light-matter interactions occur when an electromagnetic wave propagates through biological tissue. All the relevant optical properties of dielectric materials can be described by their relative permittivity ε_r, The polarization P of the material, caused by bound charges and/or polar molecules, represents the collective dipole moment per unit volume of the medium. The polarization therefore describes the interaction between the applied electric field and the medium, and is expressed as

$$P = np = \chi \varepsilon_0 E, \tag{12.9}$$

where n is the volume density of dipoles, p is their dipole moment, χ is the electric susceptibility, ε_0 is the dielectric constant of free space, and E is the applied electric field. According to its microscopic definition (i.e., $P = np$), the polarization depends on the density of dipoles; thus, a localized density fluctuation results in localized changes in the polarization or the dielectric constant (or refractive index), which in turn relate to macroscopic phenomena through $P = \chi \varepsilon_0 E$. Therefore, optical scattering can be considered to result from in homogeneities in the refractive index of the medium.

The propagation of light in a dielectric medium can be described using Maxwell's equations as follows:

$$\nabla \times H = J_b + \frac{\partial (\varepsilon_0 E)}{\partial t}, \tag{12.10}$$

$$\nabla \times E = -\frac{\partial B}{\partial t}, \tag{12.11}$$

$$\nabla \cdot (\varepsilon_0 E) = \rho_b, \tag{12.12}$$

and

$$\nabla \cdot B = 0, \tag{12.13}$$

where the incident or applied field induces a dipole moment by the displacement of the centers of positive and negative charges in the atom, and the collective effect of these dipole moments is the polarization. This polarization effect generates a bound charge density or bound current density; hence, Eqs. (12.10) and (12.12) can be modified by including the bound current density due to the polarization, $J_b = \frac{\partial P}{\partial t}$, and the bound (polarization) charge density, $\rho_b = -\nabla \cdot P$, as

$$\nabla \times H = \frac{\partial P}{\partial t} + \frac{\partial (\varepsilon_0 E)}{\partial t} = \frac{\partial (\varepsilon_0 E + P)}{\partial t} \tag{12.14}$$

and

$$\nabla \cdot (\varepsilon_0 E) - (-\nabla \cdot P) = \nabla \cdot (\varepsilon_0 E + P)\,. \tag{12.15}$$

Thus, using Eqs. (12.14) and (12.15), Maxwell's equations can be modified to include this effect and the relative permittivity ε_r can be derived depending on the materials, i.e.,

$$D = \varepsilon_0 E + P = (1 + \chi)\,\varepsilon_0 E = \varepsilon_r \varepsilon_0 E. \tag{12.16}$$

The wave equation in a dielectric medium with very small non-zero conductivity ($\sigma \neq 0$) is given by

$$\nabla^2 E = \mu\sigma \frac{\partial E}{\partial t} + \varepsilon\mu \frac{\partial^2 E}{\partial t^2} \tag{12.17}$$

and the propagation k can be expressed as

$$\begin{aligned} k &= \left(\omega^2 \mu\varepsilon + i\mu\sigma\omega\right)^{0.5} = \omega\sqrt{\mu\varepsilon}\left(1 + \frac{i\mu\sigma\omega}{\omega^2\mu\varepsilon}\right)^{0..5} \\ &= \omega\sqrt{\mu\varepsilon}\left(1 + i\frac{\sigma}{\omega\varepsilon}\right)^{0.5} \sim \omega\sqrt{\mu\varepsilon}\left(1 + i\frac{\sigma}{2\omega\varepsilon}\right) = k_1 + ik_2. \end{aligned} \tag{12.18}$$

Using complex notation for the susceptibility (i.e., $\chi = \chi_1 + i\chi_2$), relative permittivity ($\varepsilon_r = \varepsilon_1 + i\varepsilon_2$), and refractive index ($n = \sqrt{\varepsilon_r} = n_1 + in_2$), the corresponding field can be written as

$$E = E_0 e^{i(kz-\omega t)} = E_0 e^{i((k_1+ik_2)z-\omega t)} = E_0 e^{-k_2 z} e^{i(k_1 z-\omega t)}, \tag{12.19}$$

where the propagation constant is defined as $k = n\frac{\omega}{c} = (n_1 + in_2)\,\frac{\omega}{c} = k_1 + ik_2$. Here, it is clear that the imaginary term of the propagation constant, $k_2 = \frac{\sigma}{2}\sqrt{\frac{\mu}{\varepsilon}}$, originates from the very small non-zero conductivity σ, which contributes to the attenuation.

12.5.2 Optical Properties of Matter

Furthermore, the comparison of the forces due to electromagnetic fields shows that the electric field is the dominant mechanism responsible for the interaction between the light and the medium, as expressed by the following relationship:

$$\frac{F_H}{F_E} = \frac{qvB}{qE} = v\frac{B}{E} = \frac{v}{c} < 1. \tag{12.20}$$

Sources in the near-infrared (NIR) range are desirable for use with biological tissue, which is mostly (∼75 %) composed of water, because of the relatively low absorption and scattering characteristics compared with those in the visible or the mid- and far-infrared (FIR) spectral ranges [16]. The shorter-wavelength visible light is either highly scattered or absorbed in the tissue, and the mid- and far-IR radiation experiences large absorption by the water in the tissues. The wavelength dependence of the scattering can be explained considering the fact that the displacement of an oscillating dipole (given by $r = r_0 e^{i\omega t}$) has a linear relationship with electric field-induced acceleration a, which is proportional to the square of the frequency ω (i.e., $a = d^2r/dt^2 = -\omega^2 r_0 e^{i\omega t}$). Because the radiated power is related to the square of the electric field, it becomes proportional to the fourth power of the frequency (i.e., $\sim \omega^4$ or $\sim \frac{1}{\lambda^4}$), so that at visible wavelengths, the losses in the medium are dominated by scattering. Therefore, shorter wavelengths have larger scattering-related attenuation, which further limits the penetration depth in OCT imaging.

Photon transport in turbid tissue can be modeled numerically using Monte Carlo simulations and analytically using the radiative transfer equation (RTE), which can be simplified using the photon diffusion approximation [17]. Here, the behavior of photons is dominated by scattering, and the individual path of a given photon is modeled as a random walk so that a large number of photons exhibit diffusion in the medium [18]. OCT imaging is based on the detection of coherent photons that travel through a scattering turbid medium. In this medium, most of the incident photons experience random scattering and absorption whereas a few photons can diffuse in a straight line for short distances. These coherent photons are termed as ballistic photons, and photons which retain some degree of coherence (i.e., are slightly scattered) are referred to as snake photons [19].

12.6 Conclusions

In summary, optical coherence tomography is a powerful emerging medical diagnostic modality for minimally invasive tomographic imaging. OCT can provide images of micrometer-scale resolution in 2-D planes and 3-D volumes for various biomedical applications and fields of studies, such as ophthalmology. Image acquisition is achieved by detecting light with an instrument called optical interferometer and the tomographic image resolution is determined by the coherence length of the utilized light source. A simple line image can be reconstructed using either conventional time-domain analysis or Fourier-transformed data for resolving the depth information directly from the interferogram results. In addition, considering the characteristics and behavior of light, the NIR regime is the optimal spectral window for better performance considering the wavelength dependency of light scattering and attenuation in turbid tissues. Based on the above principles, an OCT system can be efficiently realized and operated with better performance and quality.

Acknowledgement This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (2013R1A1A2062448). The author acknowledges the partial use of contents and texts from the author's doctoral dissertation.

References

1. Boppart SA, Bouma BE, Pitris C, Tearney GJ, Fujimoto JG, Brezinski ME (1997) Forward-imaging instrumentations for optical coherence tomography. Opt Lett 22:1618–1620
2. Hartl I, Li XD, Chudoba C, Ghanta RK, Ko TH, Fujimoto JG, Ranka JK, Windeler PS (2001) Ultrahigh-resolution optical coherence tomography using continuum generation in an air-silica microsctructure optical fiber. Opt Lett 26:608–610
3. Teramura Y, Suekuni M, Kannari F (2000) Two-dimensional optical coherence tomography using spectral domain interferometry. J Opt A 2:21–26
4. Hoeling B, Fernandez A, Haskell R, Huang E, Myers W, Petersen D, Ungersma S, Wang R, Williams M, Fraser S (2000) An optical coherence microscope for 3-dimensional imaging in developmental biology. Opt Express 6:136–146
5. Xi C, Marks D, Schlachter S, Luo W, Boppart SA (2006) High-resolution three-dimensional imaging of biofilm development using optical coherence tomography. J Biomed Opt 11:034001
6. Cheng Y, Larin KV (2007) In vivo two- and three-dimensional imaging of artificial and real fingerprints with optical coherence tomography. IEEE Photon Technol Lett 19:1634–1636
7. Fujimoto JG, De Silvestri S, Ippen EP, Puliafito CA, Margolis R, Oseroff A (1986) Femtosecond optical ranging in biological systems. Opt Lett 11:150–152
8. Fercher AF, Mengedoht K, Werner W (1988) Eye length measurement by interferometry with partially coherent light. Opt Lett 13:186–188
9. Bezerra HG, Costa MA, Guagliumi G, Rollins AM, Simon DI (2009) Intracoronary optical coherence tomography: a comprehensive review. J Am Coll Cardiol Intv 2:1035–1046
10. Tasi T-H, Fujimoto JG, Mashimo H (2014) Endoscopic optical coherence tomography for clinical gastroenterology. Diagnostics 4:57–93
11. Kharchenko S, Adamowicz J, Wojtkowski M, Drewa T (2013) Optical coherence tomography diagnostics for onco-urology. Review of clinical perspectives. Cen Eur J Urol 66:136–141
12. Babalola O, Mamalis A, Lev-Tov H, Jagdeo J (2014) Optical coherence tomography (OCT) of collagen in normal skin and skin fibrosis. Arch Dermatol Res 306:1–9
13. Kang JU, Huang Y, Zhang K, Ibrahim Z, Cha J, Lee WPA, Brandacher G, Gehlbach PL (2012) Real-time three-dimensional Fourier-domain optical coherence tomography video image guided microsurgeries. J Biomed Opt 17:081403
14. Pedrotti FL, Pedtrotti LS (1993) Introduction to optics, 2nd edn. Prentice Hall, Englewood Cliffs
15. Schmitt JM (1999) Optical coherence tomography (OCT): a review. IEEE J Quantum Electron 5:1205–1215
16. Jacues SL et al (1987) Angular dependence of HeNe laser light scattering by human dermis. Lasers Life Sci I:309–334
17. Haskell RC, Svaasand LO, Tsay T-T, Feng T-C, McAdams MS, Tromberg BJ (1994) Boundary conditions for the diffusion equation in radiative transfer. J Opt Soc Am 11:2727–2741
18. Yoo KM, Alfano RR (1990) Time-resolved coherent and incoherent components of forward light scattering in random media. Opt Lett 15:320–322
19. Farsiu S, Christofferson J, Eriksson B, Milanfar P, Friedlander B, Shakouri A, Nowak R (2007) Statistical detection and imaging of objects hidden in turbid media using ballistic photons. Appl Opt 46:5805–5822

Chapter 13
Detection of Brain Metastases Using Magnetic Resonance Imaging

Jaeseok Park

Abstract In detecting brain metastases, three-dimensional (3D) magnetic resonance imaging (MRI), which exploits a T_1-weighted contrast mechanism, has been widely used after administering T_1-enhancing contrast agents. However, since contrast materials remain in both blood and the tumor parenchyma and increases the signal intensity of both regions, it is often challenging to differentiate brain tumors from blood. The purpose of this work is to develop a novel, highly selective whole-brain metastases MRI, wherein the signal intensity of the brain tumor is enhanced while that of blood is suppressed. A 3D non-CPMG fast/turbo spin echo pulse sequence, which incorporates variable refocusing-flip-angles and flow-sensitizing gradients, was employed to suppress blood signals. To avoid loss of signals in stationary tissues resulting from the non-CPMG condition, the first refocusing-flip-angle was forced to 180°. Simulations and in vivo volunteer and patient experiments were performed to demonstrate the effectiveness of this approach in detecting small brain metastases.

Keywords Magnetic resonance imaging • Brain metastases • Black-blood • Fast spin echo • Contrast-enhanced

13.1 Introduction

In brain metastasis, blood vessels do not maintain the blood-brain-barrier. Contrast agents with small molecular weights, if administered exogenously, accumulate within the extravascular extracellular space. Thus, contrast-enhanced (CE) imaging methods [1–6], which are sensitive to the T_1-shortening effect of contrast materials, have been widely used for the detection of brain metastases. Among them, magnetization-prepared-rapid-gradient-echo (MP-RAGE) imaging [7, 8], a three-dimensional (3D) T_1-weighted imaging sequence, has been used to acquire high-resolution, isotropic whole-brain data in a clinically acceptable imaging time.

J. Park (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: jaeseokpark.biel@gmail.com

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_13

Despite the advantages, a potential confound is the distribution of contrast agents within blood vessels and the tumor parenchyma, resulting in the simultaneous enhancement of both regions and thereby impairing the accuracy of metastatic lesion detection. Since blood vessels close to brain tumors may result in diagnostic confusion, it is necessary to selectively enhance the signal intensity of the brain tumors while suppressing the blood vessels elsewhere in the brain.

Black-blood imaging methods [9–14], if successfully combined with CE 3D MP-RAGE, may have a potential to differentiate brain tumours from blood vessels. Nevertheless, it is challenging to achieve CE black-blood imaging using MP-RAGE due to rapid recovery of blood signals along the echo train. Furthermore, blood suppression becomes more problematic with increasing spatial coverages. Additionally, motion-sensitizing magnetization preparation, which employs a 90° (excitation pulse) – 180° (refocusing pulse) – 90° (flip-back restore pulse) pulse scheme with matching crushers inserted on either side of the refocusing pulse, may yield substantial T_2-weighting, reducing the signal intensity of CE tumors.

In this work, we develop a novel, T_1-weighted, black-blood version of the single-slab 3D turbo/fast SE pulse-sequence, investigating its feasibility as a novel CE whole-brain black-blood imaging for efficiently detecting small metastases. Numerical simulations and experimental studies were performed in volunteers and patients at 3T using conventional 3D MP-RAGE and the proposed method for comparison.

13.2 Method

13.2.1 T_1-Weighted, Black-Blood Single-Slab 3D Turbo/Fast SE

A schematic and timing diagram of the proposed T_1-weighted single-slab 3D non-CPMG turbo/fast SE pulse sequence is shown in Fig. 13.1. A spatially non-selective excitation radio-frequency (RF) pulse is applied for the selection of the whole-brain. Short, non-selective RF pulses (duration, 400 μs) are employed in the refocusing pulse train, which shortens echo spacing (ESP) (time between neighboring echoes) and thus enhances the efficiency of data acquisition. To reduce power deposition to human tissues, permit a long echo train (a series of RF pulses), and maintain T_1-weighted signal evolution along the entire echo train, variable flip angles [15, 16] calculated using gray matter (GM) specific signal prescription are used in the refocusing pulse train. The refocusing flip angles typically vary rapidly from high to low values in the beginning of the echo train to store magnetization along the longitudinal axis and contribute to signals later in the echo train, and then increase slowly to counter an inherent signal loss due to relaxation. At the end of the refocusing pulse train, composite restore RF pulses composed of three short, non-selective pulses are applied to prepare partial inversion recovery for next excitation time of repetition.

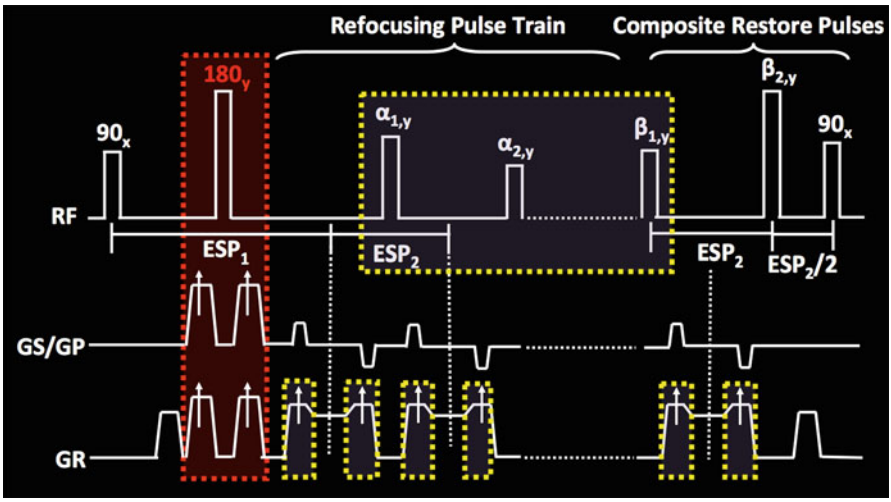

Fig. 13.1 A schematic and timing diagram of the proposed pulse sequence for the detection of brain metastases

In the frequency encoding (FE) direction in conventional turbo/fast SE pulse-sequence, the magnetization of flowing spins in the SE pathway experiences 90°–180°–180°–like RF pulses and thus remains in the transverse plane during all the segments of the pulse sequence, undergoing flow dephasing at every odd echo and flow compensation at every even echo. On the other hand, magnetizations in the stimulated echo (STE) pathway see 90°–90°–90°–like RF pulses and thus experiences flow dephasing throughout the echo train. For flowing blood, SE and STE signal may re-phase at different time points during data acquisition, yielding loss of signals resulting from phase cancellations. Flow-induced phase-dispersion would be further activated with increasing incoherence of SE and STE. Given the consideration above, variable-flip-angle approach in the refocusing pulse train is advantageous in that the initial low flip angles store substantial magnetizations along the longitudinal axis and thereby increase contribution of STE signals later in the echo train. To enlarge phase dispersions for magnetizations in the STE pathway and those in the SE pathway, the gradient moment of matching crushers (spoilers) on either side of each refocusing pulse in the FE direction is increased. Additionally, to decrease blood signals in the beginning of the echo train, additional matching crushers are inserted on either side of the first refocusing pulse in the x, y, and z-axes. Given the desired gradient moment, all matching crushers are designed to have the shortest possible duration within the maximum gradient amplitude and slew rate. To retain high echo-train efficiency, a longer ESP (ESP_1), which accommodates the additional matching crushers on either side of the first refocusing pulse, is used for the first echo, while the shortest possible ESP (ESP_2) is employed for the remainder of the echo train. The non-CPMG configuration ($ESP_1 \neq ESP_2$) does not re-phase STE signals in a coherent position with SE signals, resulting in loss of signals in

stationary tissues. This problem is addressed in this work by forcing the flip angle of the first refocusing RF pulse to 180°. Data acquisition is skipped for the first two echoes to achieve substantial blood suppression.

13.2.2 Numerical Simulation

Numerical simulation of the Bloch-equation is performed under the assumption that: (1) all RF pulses are short, non-spatially-selective, (2) gradient pulses use rectangular waveforms and are applied in the FE direction, wherein the duration of the pre-phasing gradient immediately after the excitation pulse is ESP/2, that of the additional matching crushers on either side of the first refocusing pulse is ESP/2, and that of the re-phasing gradient between a pair of non-spatially-selective refocusing pulses is ESP, and (3) blood flows in parallel with the FE direction at a constant velocity. The re-phasing gradient was simulated by successively applying the pre-phasing gradient twice.

Since all RF pulses are very short and thus relaxation during pulse application is negligible, the effect of each RF pulse on magnetization is simply a rotation about an axis in the transverse plane. The signal intensity averaged over spin isochromats at each time of echo is calculated by*:

$$S(n) = \frac{1}{N}\sum_{l=1}^{N}\left(\prod_{m=1}^{n} M(l,m)\, R_x\left(90^\circ\right) M_i\right) \tag{13.1}$$

$$M(l,m) = T(\tau(m))\, R_z\left(\phi_s(l,m) + \phi'_M(m)\right) R_y\left(\alpha_m^0\right) T(\tau(m))\, R_z\left(\phi_S(l,m) + \phi_M(m)\right) \tag{13.2}$$

In the presence of the additional matching crushers on either side of the first refocusing pulse,

$$\tau(m) = \begin{cases} ESP, & m = 1 \\ ESP/2, & m > 1 \end{cases} \tag{13.3}$$

$$\phi_S(l,m) = \begin{cases} \left(1 + \frac{G_D}{G_R}\right)\frac{360^\circ}{N}(l-1), & m = 1 \\ \frac{360^\circ}{N}(l-1), & m > 1 \end{cases} \tag{13.4}$$

$$\phi_M(m) = \begin{cases} \gamma v M_1 + 3\frac{G_D}{G_R}\gamma v M_1, & m = 1 \\ (4m+1)\gamma v M_1, & m > 1 \end{cases} \tag{13.5}$$

$$\phi'_M(m) = \begin{cases} 5\frac{G_D}{G_R}\gamma v M_1 + 7\gamma v M_1, & m = 1 \\ (4m+3)\gamma v M_1, & m > 1 \end{cases} \tag{13.6}$$

In the absence of the additional matching crushers on either side of the first refocusing pulse,

$$\tau(m) = ESP/2,\ m \geq 1 \tag{13.7}$$

$$\phi_S(l,m) = \tfrac{360^\circ}{N}(l-1),\ m \geq 1 \tag{13.8}$$

$$\phi_M(m) = (4m-3)\,\gamma v M_1,\ m \geq 1 \tag{13.9}$$

$$\phi_M(m) = (4m-1)\,\gamma v M_1,\ m \geq 1 \tag{13.10}$$

$$M_1 = \frac{1}{2} G_R \left(\frac{ESP}{2}\right)^2 \tag{13.11}$$

where n is the echo number, N is the number of isochromats, l is the isochromat index, m is the refocusing pulse number, α^o is the refocusing flip angle, $M(l,m)$ is the magnetization vector as a function of l and m, $R_x(\ldots)$, $R_y(\ldots)$, and $R_z(\ldots)$ are the rotation matrices about x, y, and z-axes, respectively, M_i is the initial magnetization vector, $[0\ 0\ M_0]^T$, $T(\ldots)$ is the matrix operator that applies T_1 and T_2 relaxation for the specified time, ϕ_S is the gradient-induced phase for stationary tissues, ϕ_M and $\phi_M^{'}$ are the flow-induced phase for the 1st order gradient moment, G_R is the amplitude of the pre-phasing gradient, G_D is the amplitude of the additional matching crushers on either side of the first refocusing pulse, γ is the gyromagnetic ratio, v is the constant flow velocity, and M_1 is the first-order moment of the pre-phasing gradient. The following parameters are retained throughout the simulations: ESP = 4.0 ms; T_1 and T_2 relaxation values for CE blood = 50 ms and 40 ms, respectively; and ETL = 22.

To investigate the effect of increasing magnetizations in the STE pathway on CE blood suppression, numerical simulations are performed for four different refocusing-flip-angle schemes (constant flip angle (CFA) = 180°; CFA = 120°; variable FA (VFA), $\alpha_1 \neq 180°$; VFA, $\alpha_1 = 180°$) with the following parameters: $G_R = 10$ mT/m and $G_D = 0$ mT/m. The effect of flow velocity on the signal intensity of flowing blood is evaluated at $v = 5$ cm/s.

13.2.3 Experimental Studies

Imaging studies were performed in volunteers and patients suspected of brain metastasis before and after the administration of a double dose of the contrast agent at a 3T whole-body MR scanner (MAGNETOM Trio, Siemens Medical Solutions, Erlangen, Germany). Informed written consent was obtained prior to imaging.

To compare the proposed method with conventional MP-RAGE in detecting small metastatic brain tumors, two sets of high-resolution isotropic whole brain images in the sagittal plane were acquired after the administration of contrast agent. The imaging parameters common to both methods were: FOV = 250 × 250 mm^2; partitions = 176; slice thickness = 1.0 mm; and imaging time = 8.5 min. The imaging parameters specific to MP-RAGE were: TR/TE/TI = 2200/4.4/900 ms; in-plane acquisition matrix = 256 × 256; and flip angle = 10°; while those specific for the proposed method were: TR/TE$_{eff}$ = 650/11.7 ms; in-plane acquisition matrix = 240 × 256; ESP$_1$/ESP$_2$ = 18.9/4.3 ms; ETL = 22; VFA ($\alpha_1 = 180°$); and $R_g = 2.5$; $R_d = 2.0$–3.0.

13.3 Results

13.3.1 Numerical Simulation

Figure 13.2 represents the effect of the refocusing-flip-angle schemes in the proposed method on CE blood (flowing at 5 cm/s). In the CFA scheme, the signal intensity in CE blood decreases with decreasing flip angles from 180° to 120°. Unlike CFA, VFA exhibits much lower blood signal intensity, particularly in the first, several echoes. Additionally, VFA scheme with $\alpha_1 = 180°$ appears to yield signal evolution similar but shifted by one echo as compared to that with $\alpha_1 \neq 180°$.

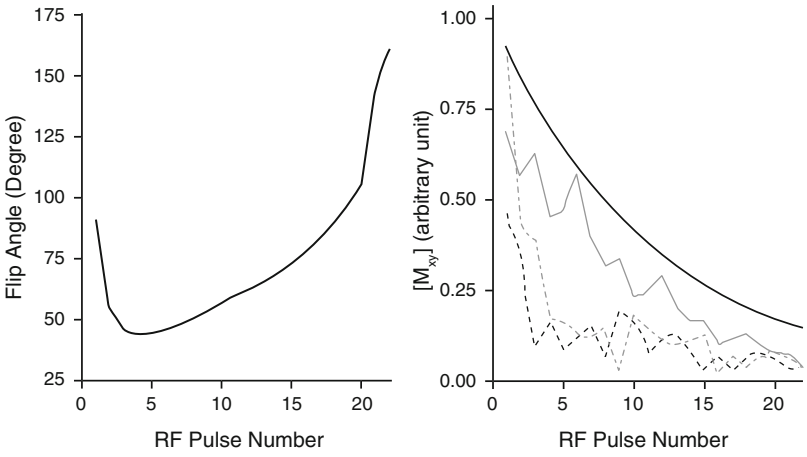

Fig. 13.2 Variable refocusing flip angles in the proposed pulse sequence with increasing RF pulse number (*left*) and its corresponding signal evolution of flowing spins at 5 cm/s (*right*) (*heavy solid*: CFA with 180°, *light solid*: CFA with 120°, *unevenly dotted line*: VFA with the first 180° refocusing flip angle, *evenly dotted line*: VFA without the first 180° refocusing flip angle)

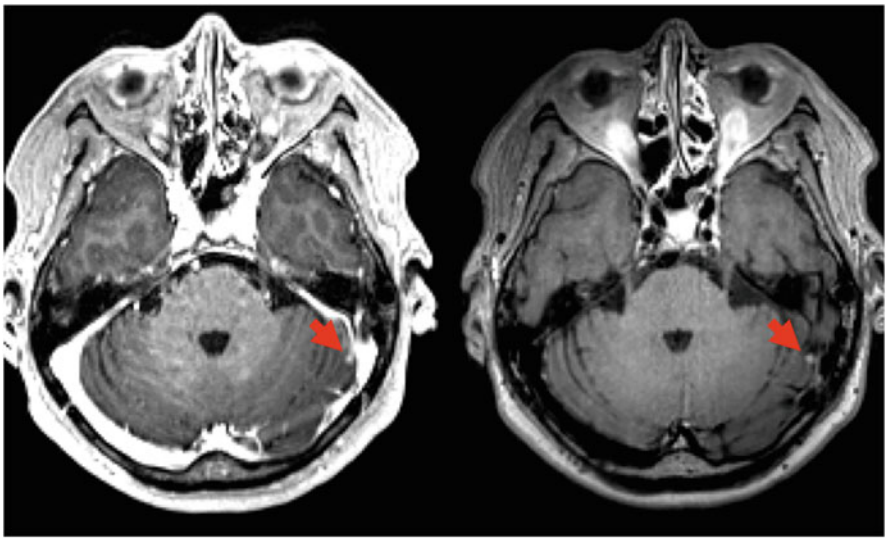

Fig. 13.3 Comparison of images acquired using conventional MP-RAGE (*left*) and the proposed method (*right*). *Arrows* represent metastatic brain tumors

13.3.2 Experimental Studies

Figure 13.3 compares the images acquired in patients using conventional MP-RAGE and the proposed method (VFA with $\alpha_1 = 180°$, $R_g/R_d = 2.5/2.0–3.0$) following the administration of contrast materials. Both metastatic brain tumors and blood in conventional MP-RAGE are simultaneously enhanced, while in the proposed method metastatic tumor signals remain high but blood signal is suppressed.

13.4 Conclusion

In conclusion, we have demonstrated the effectiveness of the proposed method over conventional MP-RAGE in detecting small brain metastases by enhancing tumor signals while selectively suppressing blood signals. Since this work is confined in investigating the technical feasibility for efficiently detecting small brain metastases, the clinical utility of the proposed method needs to be evaluated in future studies.

Acknowledgement Numerical simulation in this manuscript follows closely a prior published paper by the authors [17]

References

1. Chappell PM, Pelc NJ, Foo TK, Glover GH, Haros SP, Enzmann DR (1994) Comparison of lesion enhancement on spin-echo and gradient-echo images. AJNR Am J Neuroradiol 15(1):37–44
2. Finelli DA, Hurst GC, Gullapali RP, Bellon EM (1994) Improved contrast of enhancing brain lesions on postgadolinium, T1-weighted spin-echo images with use of magnetization transfer. Radiology 190(2):553–559
3. Sugahara T, Korogi Y, Ge Y, Shigematsu Y, Liang L, Yoshizumi K, Kitajima M, Takahashi M (1999) Contrast enhancement of intracranial lesions: conventional T1-weighted spin-echo versus fast spin-echo MR imaging techniques. AJNR Am J Neuroradiol 20(8):1554–1559
4. Li D, Haacke EM, Tarr RW, Venkatesan R, Lin W, Wielopolski P (1996) Magnetic resonance imaging of the brain with gadopentetate dimeglumine-DTPA: comparison of T1-weighted spin-echo and 3D gradient-echo sequences. J Magn Reson Imaging 6(3):415–424
5. Mirowitz SA (1992) Intracranial lesion enhancement with gadolinium: T1-weighted spin-echo versus three-dimensional Fourier transform gradient-echo MR imaging. Radiology 185(2):529–534
6. Walker MT, Kapoor V (2007) Neuroimaging of parenchymal brain metastases. Cancer Treat Res 136:31–51
7. Brant-Zawadzki M, Gillan GD, Nitz WR (1992) MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. Radiology 182(3):769–775
8. Mugler JP 3rd, Brookeman JR (1990) Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magn Reson Med 15(1):152–157
9. Felmlee JP, Ehman RL (1987) Spatial presaturation: a method for suppressing flow artifacts and improving depiction of vascular anatomy in MR imaging. Radiology 164(2):559–564
10. Ehman RL, Felmlee JP (1990) Flow artifact reduction in MRI: a review of the roles of gradient moment nulling and spatial presaturation. Magn Reson Med 14(2):293–307
11. Edelman RR, Chien D, Kim D (1991) Fast selective black blood MR imaging. Radiology 181(3):655–660
12. Chien D, Goldmann A, Edelman RR (1992) High-speed black blood imaging of vessel stenosis in the presence of pulsatile flow. J Magn Reson Imaging 2(4):437–441
13. Nguyen TD, de Rochefort L, Spincemaille P, Cham MD, Weinsaft JW, Prince MR, Wang Y (2008) Effective motion-sensitizing magnetization preparation for black blood magnetic resonance imaging of the heart. J Magn Reson Imaging 28(5):1092–1100
14. Wang J, Yarnykh VL, Hatsukami T, Chu B, Balu N, Yuan C (2007) Improved suppression of plaque-mimicking artifacts in black-blood carotid atherosclerosis imaging using a multislice motion-sensitized driven-equilibrium (MSDE) turbo spin-echo (TSE) sequence. Magn Reson Med 58(5):973–981
15. Park J, Mugler JP 3rd, Horger W, Kiefer B (2007) Optimized T1-weighted contrast for single-slab 3D turbo spin-echo imaging with long echo trains: application to whole-brain imaging. Magn Reson Med 58(5):982–992
16. Mugler JP, 3rd (2003) Practical implementation of optimized tissue-specific prescribed signal evolutions for improved turbo-spin-echo imaging. In: Proceedings of the 11th annual meeting of ISMRM, Toronto, Canada, p 203
17. Park J, Kim EY (2010) Contrast-enhanced three-dimensional whole-brain black-blood imaging: application to brain metastases. Magn Reson Med 63(3):553–561

Chapter 14
Deep Learning in Diagnosis of Brain Disorders

Heung-Il Suk, Dinggang Shen, and the Alzheimer's Disease Neuroimaging Initiative*

Abstract In this chapter, we introduce our recent work on neuroimaging-based AD diagnosis with machine learning techniques, especially deep learning. Specifically, we focus on the problems of feature representation and complementary information fusion from different modalities, e.g., MRI and PET. In our experimental results on the publicly available ADNI dataset, we could validate the effectiveness of the deep learning-based feature representation and its superiority to the competing methods. We also present the importance of collaborating communities of machine learning and clinical neuroscience for clinical interpretation of the learned feature representations.

Keywords Alzheimer's disease (AD) • Mild cognitive impairment (MCI) • Deep learning • Stacked auto-encoder • Deep Boltzmann machine

14.1 Introduction

As the population becomes older, the world is now facing an epidemic of dementia. Among various causes of dementia, Alzheimer's Disease (AD) is the most prevalent in elderly people, which rises significantly every year in terms of the proportion of

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete list of ADNI investigators is available at http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Authorship_List.pdf.

H.-I. Suk (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: hisuk@korea.ac.kr

D. Shen
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: dgshen@med.unc.edu

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_14

cause of death. A recent study by Alzheimer's Association reported that 10∼20 % of people aged 65 or older have Mild Cognitive Impairment (MCI), a prodromal stage of AD [1]. But there is no treatment to halt its progression to AD yet. In this regard, it has been one of the major issues to understand the underlying mechanisms that develop such devastative neurodegenerative disease in the fields of neuroscience, neuropsychiatry, etc.

The current scientific technologies of medical imaging, such as Magnetic Resonance Imaging (MRI) and Positron Emission Topography (PET), provide paths to investigate the structure and function of the brain in vivo. With the help of such tools, researchers have made a great leap in understanding the disease. However, the group-level analysis prevalently used for investigation and understanding of the disease is not clinically applicable for individual diagnosis. In the meantime, machine learning techniques, which can efficiently analyze the complex patterns in observations, help pave the way for a computer-aided AD diagnosis system by building computational models that can discriminate patients with the disease from healthy normal subjects.

The conventional computer-aided diagnostic systems mostly considered neuroimaging features such as voxel intensities of predefined regions, gray matter volumes, cortical thickness, to name a few, all of which can be considered as simple low-level features. From a machine learning point of view, it is beneficiary to exploit the latent high-level features inherent in data to enhance the diagnostic performance. Deep learning [3], which has already proved its effectiveness by showing promising results in various fields including speech/object recognition [9, 23] and medical imaging analysis [14, 26], can discover latent or abstract high-level information in neuroimaging data, and thus be useful for the disease diagnosis. In this chapter, we introduce our recent work on neuroimaging-based AD diagnosis with deep learning.

14.2 Background

Figure 14.1 illustrates the general framework of machine learning-based AD diagnosis, composed of four main steps, namely, (1) neuroimaging data acquisition, (2) image preprocessing (including registration, tissue segmentation, Regions-

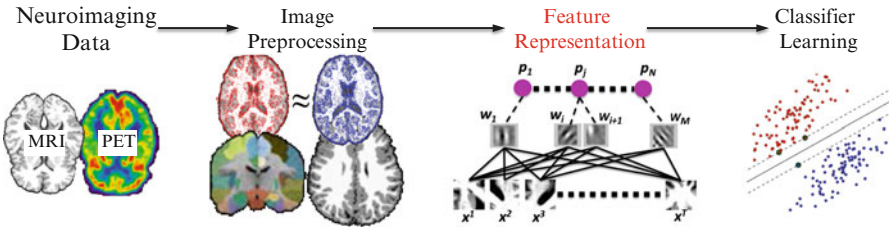

Fig. 14.1 A general framework for machine learning-based AD diagnosis using neuroimaging data

Of-Interest (ROIs) parcellation, etc.), (3) feature extraction/representation, and (4) classifier learning. Although machine learning techniques can be involved in all of these steps [11, 24], in this chapter, we focus on the step of feature representation.

As for feature extraction or representation, the existing methods can be categorized into voxel-based approach, ROI-based approach, and patch-based approach. A voxel-based approach directly uses the voxel intensities of MRI or PET as features in classification [2, 10]. Although the voxel-based approach can reflect small changes in structure or function and it is easy to interpret the results, its main limitation comes from the high-dimensionality of feature vectors and also no consideration of the inter-region relation information. On the other hand, the ROI-based approach can handle the issue of high-dimensionality by extracting representative features from the structurally or functionally parcellated brain regions. Thanks to the relatively low feature-dimension and the whole brain coverage, this approach has been most widely used in the literature [6, 12, 20, 29, 32, 34]. However, the features extracted from ROIs are very coarse in the sense that they cannot reflect small or subtle changes involved in the brain diseases. In the meantime, a patch-based approach dissects a brain into small 3D patches from which it extracts features and trains a classifier for each patch location, and then combines classifiers' outputs in a hierarchical manner [15, 31]. The patch-based approach has the advantages of (1) reflecting subtle changes by using voxel-wise features as the voxel-based approach does and (2) also considering a whole brain information as the ROI-based approach does by hierarchically integrating regional information.

14.3 Deep Learning for AD Diagnosis

Although the existing methods described in Sect. 14.2 have shown their effectiveness for AD diagnosis in the literature, they mostly used the simple low-level features without considering the high-level information latent in those features. Inspired from the biological model of the human visual cortex [7, 25], recent studies in machine learning have shown that a deep architecture composed of multiple non-linear transformations is useful to find highly non-linear and complex patterns in the data [3, 18]. This motivated us to apply deep learning techniques to neuroimaging-based AD diagnosis in [30, 31], where we used Stacked Auto-Encoder (SAE) and Deep Boltzmann Machine (DBM), respectively. In the following, we introduce these studies and further discuss the future research issues that should be tackled for clinical interpretation.

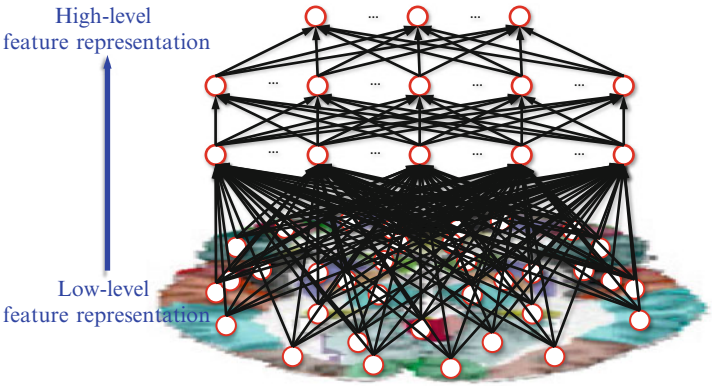

Fig. 14.2 Stacked auto-encoder that discovers the latent high-level information inherent in ROI-based features

14.3.1 Stacked Auto-Encoder (SAE)

As a pioneering study of the application of deep learning for AD diagnosis, we used an SAE [4] to discover a latent feature representation in neuroimaging or biological data. Specifically, as the name says, we stacked auto-encoders, one after another by taking the outputs from the hidden units of the lower layer as the input to the upper layer's input units, and so on. Figure 14.2 shows the network structure of our SAE model with three auto-encoders stacked. As illustrated in Fig. 14.2, in this study, we took an ROI-based approach by extracting representative features from ROIs, which set the values of the input units in the bottom layer of our SAE.

Thanks to the hierarchical nature in structure, one of the most important characteristics of the SAE is to learn or discover highly non-linear and complicated patterns such as the relations among input features. Another important characteristic of the SAE is that the latent representation can be learned directly from the data. Utilizing its representational and self-taught learning properties, we could find a latent representation of the original low-level features, directly extracted from neuroimaging data. When an input sample is presented to an SAE model, the different layers of the network represent different levels of information. That is, the lower the layer in the network, the simpler patterns (e.g., linear relations of features); the higher the layer, the more complicated or abstract patterns inherent in the input feature vector (e.g., non-linear relations among features) [17].

To find the optimal parameters, we performed unsupervised layer-wise pre-training [8] and supervised fine-tuning during the auto-encoding task via back-propagation [5, 13] sequentially. It is noteworthy that, in order to obtain the complicated non-linear relations among neuroimaging features, we considered a number of hidden units larger than the number of input features, from which we can

still find an interesting structure by imposing a sparsity constraint via a Kullback-Leibler (KL) divergence. Specifically, in our pre-training step, we optimized the following objective function:

$$E(\mathbf{Y}_{h-1}, \hat{\mathbf{Y}}_{h-1}) + \gamma \sum_{j=1}^{D_h} KL(\rho||\widehat{\rho_j}) \tag{14.1}$$

where $E(\mathbf{Y}_{h-1}, \hat{\mathbf{Y}}_{h-1})$ denotes an error between the input $\mathbf{Y}_{h-1}$ (i.e., the output from $(h-1)$-th layer) and its reconstruction $\hat{\mathbf{Y}}_{h-1}$, D_h is the number of units in the h-th hidden layer, and γ is a control parameter. In Eq. (14.1), KL divergence controls the sparseness of the hidden units based on the average activation $\hat{\rho}_j$ of the j-th hidden unit over the training samples and the target average activation ρ.

14.3.1.1 Experiments and Performance Comparison

To validate the effectiveness of the SAE-based feature representation, we conducted experiments with ADNI dataset (available at 'http://www.loni.ucla.edu/ADNI'). Specifically, we considered the baseline MRI, 18-fluoro-deoxyglucose PET, and CerebroSpinalFluid (CSF) data acquired from 51 subjects with AD, 99 subjects with MCI (including 43 progressive MCI (pMCI) and 56 stable MCI (sMCI))[1], and 52 Healthy normal Controls (HC). Along with the neuroimaging and biological data, two types of clinical scores, Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog), were also provided for each subject.

We built one SAE model[2] per modality and concatenated the original Low-Level Features (LLF) and the SAE-learned Features (SAEF) to construct an augmented feature vector (LLF+SAEF), which thus includes both low-level and high-level information. To fuse the complementary information from multiple modalities, we used a multi-kernel Support Vector Machine (SVM) [27], preceded by feature selection with a sparse regression method [33]. We considered three binary classification problems: AD vs. HC, MCI vs. HC, and pMCI vs. sMCI. In the classification of MCI vs. HC, both pMCI and sMCI data were used as the MCI class. Due to a limited small number of training samples, we applied a 10-fold cross validation technique.

We summarized the classification accuracies in Table 14.1. In AD vs. HC, compared to the accuracy of 0.970 with an LLF-based method, the proposed method improved the accuracy by 0.009. In the classification of MCI and HC, the

[1] In our work, 'progressive' and 'stable' denote whether the subjects with MCI progressed to AD in 18 months.

[2] The number of hidden units were manually determined proportional to the input dimension. As for the sparsity target and the weighting parameter of the sparsity penalty in Eq. (14.1), we set to $\rho = 0.05$ and $\gamma = 0.01$.

Table 14.1 Performance comparison in an ROI-based feature representation. As for statistical significance, a paired *t*-test was performed (LLF: Low-Level Features; SAEF: SAE-learned Feature representations; pMCI: progressive MCI; sMCI: stable MCI)

Method	AD/HC	MCI/HC	pMCI/sMCI
LLF	0.970±0.010	0.848±0.014	0.760±0.020
LLF+SAEF	**0.979±0.007**	**0.888±0.012**	**0.779±0.027**
p-value	0.0432	2.2693e-06	0.0904

proposed method showed the best classification accuracy of 0.888. The performance improvement compared to the classification accuracy of 0.848 with the LLF-based method was 0.04. In discriminating pMCI from sMCI, our method also outperformed the LLF-based method. While the LLF-based method showed the classification accuracy of 0.760, our method achieved the classification accuracy of 0.779. Based on these results, we argue that the SAE-based feature representation helped enhance the diagnostic accuracies, justifying the importance of using high-level information latent in the observation.

14.3.2 Deep Boltzmann Machine (DBM)

While an ROI-based approach helps alleviate the high-dimensionality problem in neuroimaging pattern analysis, it fails to handle subtle changes within an ROI or across ROIs. In this regard, Liu et al. proposed a patch-based approach that can efficiently handle both the high-dimensionality problem and subtle changes in an image and gradually integrated a number of local patches of a Gray Matter (GM) density map hierarchically [15]. Although they showed the efficacy of their method for AD/MCI diagnosis, it is well known that the structural or functional images are susceptible to acquisition noise, intensity inhomogeneity, artifacts, etc. Furthermore, the raw voxel density or intensity values in a patch can be considered as low-level features that do not efficiently capture more informative high-level features. To this end, we proposed a deep learning-based high-level structural and functional feature representation from MRI and PET, respectively, for AD/MCI classification. Furthermore, for multiple modalities fusion, unlike the existing methods that first extracted features from each modality independently and then mostly combined heterogeneous features via either simple feature concatenation or kernel machines, we designed a multi-modal deep learning architecture using DBM.

A DBM is structured by stacking multiple Restricted Boltzmann Machines (RBMs) in a hierarchical manner. The rationale of using DBM for feature representation is as follows: It can learn the internal latent representations that capture non-linear complicated patterns and/or statistical relations in a hierarchical manner [3, 16]. However, unlike many other deep network models such as deep belief network [8] and SAE [26], the approximate inference procedure after the initial bottom-up

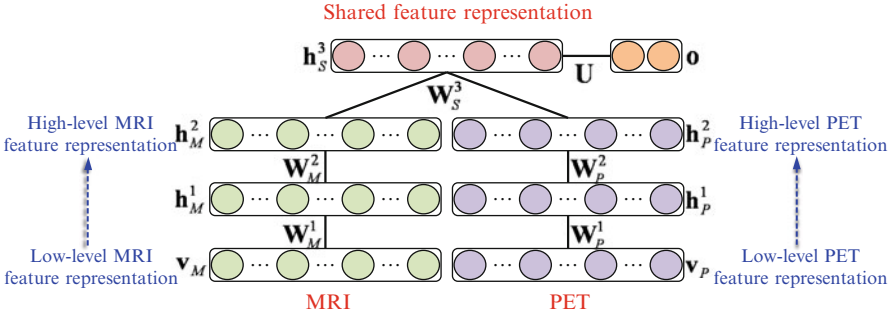

Fig. 14.3 Multimodal deep Boltzmann machine that integrates the structural and functional information and finds the shared feature representations

pass incorporates top-down feedback, which allows DBM to use higher-level knowledge to resolve uncertainty about intermediate-level features, thus creating better data-dependent representations and statistics [22]. Thanks to this two-way dependencies, i.e., bottom-up and top-down, it was shown that DBMs achieved the state-of-the-art performance in computer vision [21, 28]. To this end, we used a DBM to discover hierarchical feature representations from neuroimaging in our work.

Regarding multiple modalities fusion, different modalities will have different statistical properties. Thus, simple concatenation of the features of multiple modalities in a shallow architecture can cause strong connections among the variables of an individual modality, but failed to find inter-modality relations [19]. In order to tackle this problem, we devised a discriminative Multi-Modal DBM (MM-DBM), in which the top hidden layer had multiple entries of the lower hidden layers and the label layer, to extract a shared feature representation by fusing neuroimaging information of MRI and PET. Figure 14.3 presents a network of our MM-DBM, where one pathway represents the statistical properties of MRI and the other pathway represents those of PET, and the top shared hidden layer finally discovers the shared properties of the modalities in a supervised manner. The joint distribution over the multimodal inputs of MRI ($\mathbf{v}_M$) and PET ($\mathbf{v}_P$) and the output label ($\mathbf{o}$) can be estimated as follows:

$$P(\mathbf{v}_M, \mathbf{v}_P, \mathbf{o}; \Theta) = \sum_{\mathbf{h}_M^2, \mathbf{h}_P^2, \mathbf{h}_S} P(\mathbf{h}_M^2, \mathbf{h}_P^2, \mathbf{h}_S^3, \mathbf{o}) \left(\sum_{\mathbf{h}_M^1} P(\mathbf{v}_M, \mathbf{h}_M^1, \mathbf{h}_M^2) \right) \left(\sum_{\mathbf{h}_P^1} P(\mathbf{v}_P, \mathbf{h}_P^1, \mathbf{h}_P^2) \right) \tag{14.2}$$

where $\Theta = \{\mathbf{W}_M^1, \mathbf{W}_M^2, \mathbf{W}_P^1, \mathbf{W}_P^2, \mathbf{W}_S^3, \mathbf{U}\}$, $\mathbf{h}$ denotes a hidden layer, the subscripts M, P, and S denote, respectively, units of the MRI path, the PET path, and the shared hidden layer. For the parameters learning, we performed two consecutive steps: (1)

Table 14.2 Performance comparison in a patch-based feature representation (pMCI: progressive MCI, sMCI: stable MCI)

Method	AD/HC	MCI/HC	pMCI/sMCI
Intensity [15]	0.903±0.070	0.839±0.006	0.733±0.125
MM-DBM	**0.954±0.052**	**0.857±0.052**	**0.759±0.154**

a greedy layer-wise pre-training for a good initial setup of the model parameters and (2) iterative alternation of variational mean-field approximation to estimate the posterior probabilities of hidden units and stochastic approximation to update model parameters [22].

14.3.2.1 Experiments and Performance Comparison

We used the baseline MRI and PET data of the ADNI dataset: 93 subjects with AD, 204 subjects with MCI including 76 pMCI and 128 sMCI, and 101 HC. After conducting the preprocessing of anterior commissure-posterior commissure correction, skull stripping, cerebellum removal, registration to a common space, and tissue segmentation, we obtained spatially normalized GM volumes (i.e., GM tissue densities) and the PET images rigidly aligned to the corresponding MR images. For computational efficiency, we further down-sampled images to $64 \times 64 \times 64$ voxels. As for a patch size, we set it to $11 \times 11 \times 11$ by following Liu et al.'s work [15] for fair comparison, and thus the input dimension of each modality patch in our MM-DBM was 1,331.

We applied a 10-fold cross validation technique. The outputs of the top shared layer were used as features, which represent the fused information of structural and functional images. In order to combine the distributed patch information over an image and to build an image-level classifier, we used a hierarchical classifier learning scheme, described in [15]. In our work, we used a linear SVM for classification.

As presented in Table 14.2, in the classification of AD and HC, our method showed the mean accuracy of 0.954. Compared to the intensity-based method [15] that showed the accuracy of 0.903, we improved by 0.051. In the discrimination of MCI from HC, the proposed method achieved the accuracy of 0.857. Meanwhile, the intensity-based method [15] achieved the accuracies of 0.839. Again, the proposed method outperformed the competing method by making performance improvements of 0.018. In the classification between pMCI and sMCI, which is the most important for early diagnosis and treatment, the intensity-based method achieved the accuracy of 0.733. Compared to this result, our method improved the accuracy by 0.026. Concisely, in our three binary classifications, based on the classification accuracy, our deep learning-based method clearly outperformed the competing method.

14.4 Discussions and Conclusions

We applied deep learning methods for high-level feature representations and validated their efficacy by showing their superiority to the competing methods in terms of the diagnostic performance. Specifically, in Sect. 14.3.1, we applied an SAE to discover latent relations among the ROI-based features and then combined multiple modalities via kernel machine. Although the SAE model can be considered as the conventional multi-layer neural network, by initializing our model parameters with a greedy layer-wise pre-training and then fine-tuning the whole model, we could learn parameters to represent the inherent information better. Meanwhile, in Sect. 14.3.2, we devised a systematic method for a joint feature representation with an MM-DBM. Unlike the SAE model that learns parameters in a top-down manner, the DBM finds the optimal parameters in a bi-directional (i.e., bottom-up and top-down) manner. We utilized this favorable characteristic and successfully applied to find the shared representation from MRI and PET.

However, from a neurophysiological perspective, it is still very hard or impossible to interpret the learned representations and to understand the trained model parameters. In other words, there is no general or intuitive way to interpret the latent feature representations or the trained models. The problem of effective interpretation of the latent feature representations is a big challenge that should be tackled by the communities of machine learning and clinical neuroscience collaboratively. Furthermore, to our best knowledge, the existing disease diagnosis systems including ours output simply the clinical status of a testing subject, e.g., AD, MCI, or HC, with no presentation of the basis that supports their decision. In other words, when a subject is identified by a diagnosis system as a patient with either AD or MCI, it is clinically important to present its basis for the decision, e.g., structurally abnormal brain regions or abnormal functional connectivities observed in the neuroimaging data. Thus, all these would be our forthcoming research issues.

Acknowledgements This chapter follows closely the prior published papers [30, 31] by the authors. This work was supported in part by NIH grants EB006733, EB008374, EB009634, AG041721, MH100217, and AG042599, and partial supported by ICT R&D program of MSIP/IITP [B0101-15-0307, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)].

References

1. Alzheimer's Association (2012) 2012 Alzheimer's disease facts and figures. Alzheimer's & Dement 8(2):131–168
2. Baron J, Chételat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F (2001) In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. NeuroImage 14(2):298–309
3. Bengio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127

4. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. In: Schölkopf B, Platt J, Hoffman T (eds) Advances in neural information processing systems, vol 19. MIT Press, Cambridge, pp 153–160
5. Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, New York
6. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 32(12):2322.e19–2322.e27
7. Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biolog Cybern 36(4):93–202
8. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
9. Hinton G, Deng L, Yu D, rahman Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Dahl TSG, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process Mag 29(6):82–97
10. Ishii K, Kawachi T, Sasaki H, Kono AK, Fukuda T, Kojima Y, Mori E (2005) Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. Am J Neuroradiol 26:333–340
11. Kim M, Wu G, Wang Q, Lee SW, Shen D (2015) Improved image registration by sparse patch-based deformation estimation. NeuroImage 105:257–268
12. Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW Jr, Jack CR, Weiner MW, Thompson PM (2010) Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging 31(8):1429–1442
13. LeCun Y, Bottou L, Orr G, Müller KR (1998) Efficient backprop. In: Orr G, Müller KR (eds) Neural networks: tricks of the trade. Lecture notes in computer science, vol 1524. Springer, Berlin/Heidelberg, pp 9–50
14. Liao S, Gao Y, Oto A, Shen D (2013) Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In: Medical image computing and computer-assisted intervention (MICCAI 2013), Nagoya. Lecture notes in computer science, vol 8150, pp 254–261
15. Liu M, Zhang D, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2013) Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. Hum Brain Mapp 35(4):1305–1319
16. Mohamed A, Dahl GE, Hinton GE (2012) Acoustic modeling using deep belief networks. IEEE Trans Audio Speech Lang Process 20(1):14–22
17. Montavon G, Braun ML, Müller KR (2011) Kernel analysis of deep networks. J Mach Learn Res 12:2563–2581
18. Montavon G, Orr GB, Müller KR (eds) (2012) Neural networks: tricks of the trade. Lecture notes in computer science, vol 7700, 2nd edn. Springer, Berlin/Heidelberg
19. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY (2011) Multimodal deep learning. In: Proceedings of the 28th international conference on machine learning, Bellevue, pp 689–696
20. Nordberg A, Rinne JO, Kadir A, Langstrom B (2010) The use of PET in Alzheimer disease. Nat Rev Neurol 6(2):78–87
21. Salakhutdinov R, Hinton GE (2009) Deep Boltzmann machines. In: Proceedings of the international conference on artificial intelligence and statistics, Clearwater Beach, pp 448–455
22. Salakhutdinov R, Hinton G (2012) An efficient learning procedure for deep Boltzmann machines. Neural Comput 24(8):1967–2006
23. Salakhutdinov R, Tenenbaum J, Torralba A (2013) Learning with hierarchical-deep models. IEEE Trans Pattern Anal Mach Intell 35(8):1958–1971
24. Sanroma G, Wu G, Gao Y, Shen D (2014) Learning to rank atlases for multiple-atlas segmentation. IEEE Trans Med Imaging 33(10):1939–1953
25. Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition, San Diego, vol 2, pp 994–1000

26. Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO (2013) Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell 35(8):1930–1943
27. Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B (2006) Large scale multiple kernel learning. J Mach Learn Res 7:1531–1565
28. Srivastava N, Salakhutdinov R (2012) Multimodal learning with deep Boltzmann machines. In: Pereira F, Burges C, Bottou L, Weinberger K (eds) Advances in neural information processing systems, vol 25, Curran Associates, Inc., pp 2231–2239
29. Suk HI, Lee SW (2013) A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces. IEEE Trans Pattern Anal Mach Intell 35(2):286–299
30. Suk HI, Lee SW, Shen D (2015) Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct 220(2):841–859
31. Suk HI, Lee SW, Shen D (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 101:569–582
32. Suk HI, Lee SW, Shen D (2014) Subclass-based multi-task learning for Alzheimer's disease diagnosis. Front Aging Neurosci 6(168):1–12
33. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B 68(1):49–67
34. Zhang D, Shen D (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59(2):895–907