Analysis Methods toward Brain-Machine Interfaces in Real Environments

by

Hiroshi Morioka
森岡 博史

A dissertation submitted to the Graduate School of Informatics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Informatics

at Kyoto University

Abstract

In recent years, brain-machine interfaces (BMIs) have been the focus of attention in both biomedical and engineering research fields because BMIs have potential capability to assist handicapped people and to enhance people's lifestyles. With this growing interest in BMIs, developing new methods of decoding that are non-invasive and portable, based on such as electroencephalography (EEG) and near-infrared spectroscopy (NIRS), is of major importance. Modalities such as EEG and NIRS are the only ones that are applicable to non-laboratory based environments and suitable for practical, real-world BMI applications. However, the use of EEG and NIRS involves certain inherent problems. EEG signals are generally a mixture of neural activity from broad areas (i.e., volume conduction effects), some of which may not be related to the task targeted by BMI, hence impairing BMI performance. NIRS has an inherent time delay as it measures blood flow, which therefore detracts from practical real-time BMI usefulness. Moreover, measurement signals by these modalities suffer from variability across different subjects, or variability due to physical and mental drift in a single subject during different recording sessions. Consequently, these difficulties deteriorate the performance of BMI and decrease the information-transfer rate (ITR). Thus, it is hoped that novel methods are proposed to solve these difficulties of EEG-NIRS-based BMI in order to realize dependable BMI systems working in real environments.

This thesis illustrates three approaches which have the potential to develop BMI in real environments. Chapter 3 describes a method for cortical current estimation from EEG with the help of information from NIRS. This method is effective for studying human brain activity at the cortical level with a higher spatial resolution than that of EEG sensors, with less contamination caused by EEG volume conduction effects. It is then effective and workable in non-laboratory, naturalistic conditions. Chapter 4 presents a novel methodology to robustly analyze multi-subject brain activity with inherent variability across subjects. This method is based on the unsupervised signal processing technique of dictionary learning, which was extended to compensate for variations between subjects and sessions, and is effective for analyzing large-scale brain databases consisting of inherent variability across subjects. We applied these two novel methods to a dataset of EEG and NIRS recordings during a newly-proposed selective visual-spatial attention task from multiple subjects. We thus confirmed the effectiveness of these methods for analyzing multi-subjects' EEG-NIRS data under contamination of non-target neural activities and inherent variability across subjects. Chapter 5 describes a method of robot navigation for realizing a semi-autonomous BMI-controlled wheelchair that is usable even by subjects who have relatively low ITR. The combination of BMI with robotics can provide a promising framework by complementing the limited decoding accuracy and ITR with external autonomous devices. All of these novel methods have potential to solve important issues faced by the EEG-NIRS-based BMI in real environments, and thus deserve broad exposure within the BMI community.

Acknowledgments

My sincere thanks go to my advisor, Prof. Shin Ishii, who has given continuous guidance and encouragement. I greatly appreciate the consistent support and advice of Dr. Atsunori Kanemura. I am deeply grateful to Dr. Motoaki Kawanabe for kindly sharing his ideas. I am grateful to Prof. Shin-ichi Maeda for informative discussions. I would like to thank Prof. Shigeyuki Oba for giving advice about statistical analysis. I would like to thank Prof. Osamu Hasegawa for persistent discussions. I would like to thank Dr. Jun-ichiro Hirayama for theoretical support. I would like to express my gratitude to Dr. Okito Yamashita for giving insightful advice on my study. I am grateful to Satoshi Morimoto for technical support in the measuring and analyzing of neuroimaging data. I would like to thank Dr. Taku Yoshioka for advice on variational bayesian multimodal encephalography. I would like to offer my thanks to Dr. Takeshi Ogawa for generous support. I would like to express my appreciation to Dr. Wako Yoshida for honest discussion. I would like to offer my special thanks to Dr. Ben Seymour. I would not have been able to get my first article accepted without his advice. I would like to thank Dr. Manabu Shikauchi and Dr. Shigeyuki Ikeda for kind discussions and data acquisition. I would like to express my gratitude to Dr. Naoto Yukinawa, Dr. Naoki Honda, Dr. Motoori Kon, Dr. Ken Nakae, and Dr. Henrik Skibbe for assistance in my academic life. I thank Dr. Masaki Maruyama and Yumi Shikauchi for experimental task design. Technical assistance during the experiments by Nobuo Hiroe was invaluable. I received informative support about variational bayesian multimodal encephalography software from Ryosuke Hayashi and Takao Sako. I would like to thank Dr. Yusuke Takeda and Dr. Masaaki Sato for brain activity analysis. I wish to thank all members of the Ishii lab, DBI, and CBI. I am deeply grateful to Chihiro Uo, Akiko Fujimaki, Misa Fukuda, and Junko Iwahashi for their extensive assistance with paperwork. Special thanks to my family for their consistent support and warm encouragement.

I would like to thank the members of the thesis committee, Prof. Shin Ishii, Prof. Toshiyuki Tanaka, and Prof. Manabu Kano, for reviewing this thesis and giving me useful advice and guidance.

I acknowledge financial support from the Japan Society of the Promotion of Science, which made it possible to complete this study. This research was also supported in part by a contract entitled "Novel and innovative R&D making use of brain structures" with the Ministry of Internal Affairs and Communications, Japan, by a contract entitled "Multimodal integration for brain imaging measurements" with the National Institute of Information and Communications Technology, Japan, and by JSPS KAKENHI.

Contents

List of Figures

List of Tables

Table. 1: List of abbreviations.

Abbreviation	Definition
2D	two-dimensional
3D	three-dimensional
ARD	Automatic relevance determination
BMI	Brain-machine interface
BOLD	Blood-oxygen level dependent
CBF	Cerebral blood flow
CMS	Common mode sense
CSF	Cerebral spinal fluid
CSP	Common spatial pattern
CV	Cross-validation
DCT	Discrete cosine transform
deoxyHb	Deoxygenated hemoglobin
DOT	Diffusion optimal tomography
ECoG	Electrocorticography
EEG	Electroencephalography
EKF	Extended Kalman filter
EOG	Electrooculography
EP	Evoked potential

Table. 1: (continued)

Abbreviation	Definition
ERP	Event related potential
FEF	Frontal eye field
fMRI	Functional magnetic resonance imaging
ICA	Independent component analysis
IPS	Intraparietal sulcus
ITR	Information-transfer rate
LDA	Linear discriminant analysis
LPA	Left pre-auricular point
LRF	Laser range finder
MAP	Maximum a posteriori
MDS	Multidimensional scaling
MEG	Magnetoencephalography
MRI	Magnetic resonance imaging
MT	Middle temporal
NCV	Nested cross-validation
NIRS	Near-infrared spectroscopy
OMP	Orthogonal matching pursuit
oxyHb	Oxygenated hemoglobin
PCA	Principal component analysis
PIRF	Position invariant robust feature
pIPS	Posterior intraparietal sulcus
RANSAC	Random sample consensus
RPA	Right pre-auricular point
SIFT	Scale-invariant feature transform
SLAM	Simultaneous localization and mapping
SLR	Sparse logistic regression
SMR	Sensorimotor rhythm

Table. 1: (continued)

Abbreviation	Definition
SSVEP	Steady-state visual evoked potential
SURF	Speeded-up robust features
SVM	Support vector machine
SWF	Sliding window filter
VBMEG	Variational bayesian multimodal encephalography
VEP	Visual evoked potential
vIPS	Ventral intraparietal sulcus

Chapter 1

Introduction

Brain-machine interface (BMI) has attracted much attention in biomedical engineering fields for its usefulness in assisting handicapped people and enhancing people's lifestyles (Iturrate et al., 2009; Rebsamen et al., 2010; Tanaka et al., 2005; Wolpaw and Wolpaw, 2012). Although the brain analysis has been the source of countless research contributions, it needs a lot more efforts for realizing practical applications of them to BMI in real environments.

It is of major importance to develop new methods of neural decoding based on non-invasive and portable measurements such as electroencephalography (EEG) and near-infrared spectroscopy (NIRS) because these are the only methods that are currently applicable in non-laboratory environments and suitable for practical real-world BMI applications. Although there exist several technologies for measuring human brain activity, non-invasive methods are preferred owing to their longevity and safety, e.g., EEG, NIRS, magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Although MEG and fMRI's higher spatial/temporal resolution has led to success in existing decoding studies (Chan et al., 2011; Kamitani and Tong, 2005; Waldert et al., 2008), they lack the portability possessed by EEG and NIRS, during which the body does not need to be fixed. However, EEG and NIRS lack the decoding methodologies that can match the performance of decoding from fMRI and MEG.

The difficulties of BMIs based on the EEG and NIRS modalities mainly come from three aspects; i.e., low spatial resolution, variability across subjects, and low information-transfer rate (ITR). EEG measures the voltage fluctuations on the scalp, which result from ionic current flows caused by a large number of neurons; therefore, EEG signals comprise a mixture of signals originating from different cortical areas (i.e., volume conduction effects) (Wolters et al., 2006). The contamination of electrical measurements originating from regions not related to the task targeted by the BMI can deteriorate the BMI's performance. On the other hand, NIRS measures the hemodynamic response using near-infrared light. Since NIRS does not suffer from the volume conduction effects, it has slightly higher spatial resolution than EEG. However, NIRS has inherent time delays as it measures blood flow caused by neural metabolism, and this greatly hinders real-time BMI. Thus, analysis methods that combine the advantages of these modalities and cancel out weaknesses are desirable. The second difficulty is how to deal with undesired variability across different subjects or variability during different recording sessions from a single subject. For instance, EEG often suffers from physical and mental drift. Physical variations include misalignment of sensors (electrodes) over days or recording sessions, different shapes of the head or skull across subjects, and changes in sensor impedance over time. Even when subjects perform exactly the same task, brain activity patterns also vary substantially across subjects (Garrett et al., 2011; McIntosh et al., 2013). Brain signals vary even in the same subject because of different physical and mental conditions, or interference from task-irrelevant brain activity. Such unavoidable variability is an obstacle to the highly successful application of BMIs in daily life, and to neuroscientific group or longitudinal analyses using large-scale databases. The third difficulty comes from low ITR of BMI based on EEG and NIRS. The ITR is calculated from the number of possible choices by BMI and the probability that the desired choice will be selected (also called decoding accuracy)(Yuan et al., 2013). Although many researchers have devoted their efforts in increasing the ITR by improving decoding techniques, currently it is reaching the ceiling.

In the chapters that follow, we tackle these problems related to EEG-NIRS-based BMIs working in real environments. Chapter 2 surveys the background of BMIs. In Chapter 3, to try to improve real environment EEG-NIRS-based BMIs, we propose a novel methodology in which the subjects' mental states are decoded from cortical currents estimated from EEG, with the help of information from NIRS. Source current localization, which reconstructs cortical currents based on measurements outside the scalp, has the potential to overcome the limitations of EEG-NIRS measurements. Using a Variational Bayesian Multimodal Encephalography (VBMEG) methodology, we incorporated a novel form of NIRS-based prior-to-capture event-related desynchronization from isolated current sources on the cortical surface. Then, we applied a Bayesian logistic regression technique to decode subjects' mental states from further sparsified current sources. Applying our methodology to a spatial attention task, we found that our EEG-NIRS-based decoder exhibited significant performance improvements over decoding methods based on EEG sensor signals alone. The advancement of our methodology, decoding from current sources sparsely isolated on the cortex, was also supported by neuroscientific considerations; the intraparietal sulcus, a region known to be involved in spatial attention, was a key responsible region in our task. These results suggest that our methodology is not only a practical option for EEG-NIRS-based BMI applications, but is also a potential tool to investigate brain activity in non-laboratory and natural environments.

In Chapter 4, to try to accommodate variability across subjects, we propose 1) a method for extracting spatial bases (or a dictionary) shared by multiple subjects, by using the signal-processing technique of dictionary learning modified to compensate for variations between subjects and sessions, and 2) an approach to subject-transfer decoding that uses the resting-state activity of a previously unseen target subject as calibration data to compensate for variations, eliminating the need for a standard calibration based on task session. We applied our methodology to a dataset of EEG recordings during a selective visual-spatial attention task from multiple subjects and sessions, where the variability compensation was essential for reducing the redundancy of the dictionary, we found that the extracted common brain activities were reasonable in the light of neuroscience knowledge. The applicability to subject-transfer decoding was confirmed by improved performance over existing decoding methods. These results suggest that analyzing multi-subject brain activity on a common basis using the proposed method enables information sharing across subjects with low-burden resting calibration, and is effective for practical use of BMI in variable environments. Moreover, the novelty would contribute to the neuroscience and neuroengineering communities and deserves their broad interest because they both will be important in the coming era of large-scale brain databases in natural conditions.

In Chapter 5, to try to practically develop an application of BMI to a wheelchair which is usable even by subjects who have relatively low ITR, we propose a novel method of robot navigation to realize a stable semi-autonomous BMI-wheelchair that coexists with human beings. The proposed method is a vision-based mobile robot's simultaneous localization and mapping (SLAM) and navigation system which is applicable to the BMI-controlled wheelchair workable even in real and crowded environments. The proposed method is stable even in crowded environments by extracting robust 3D feature points from sequential vision images and odometry information. We present experiments showing the utility of our approach in crowded environments, including map building and navigation. The combination of BMI with robotics can provide a promising framework by complementing the limited decoding accuracy and ITR with external autonomous devices.

Chapter 2

Background on Brain-Machine Interfaces

A BMI is a methodology to translate neuronal signals reflecting a person's intention into commands controlling a machine (e.g., cursor, robot, and prosthesis) (Dornhege et al., 2007; Graimann et al., 2011; Tan and Anton, 2010; Wolpaw and Wolpaw, 2012). The BMI does not depend on user's muscle activity, and then can provide novel communication channels to external devices for supporting their daily life; e.g., neuroprostheses, robotic wheelchairs, mind-to-mind communication tools.

2.1 Measurement modalities

There are two strategies for recording the brains signals analyzed in BMI: invasive or noninvasive strategies. Invasive BMI, which derives the user's intention from spiking activities of cortical cells through electrode arrays implanted into the brain, has been studied mainly in non-human primates (Carmena et al., 2003; Mehring et al., 2003; Musallam et al., 2004; O'Doherty et al., 2011; Serruya et al., 2002; Taylor et al., 2002; Velliste et al., 2008; Wessberg et al., 2000). This approach is based on the finding that the direction of arm movements can be decoded from the firing rates of neurons in the motor cortex (Georgopoulos et al., 1986). Instead of implanting electrodes into the brain, electrocorticography (ECoG) is also used for recording brain activity in BMIs (Chao et al., 2010; Leuthardt et al., 2004; Mehring et al., 2004; Schalk and Leuthardt, 2011). ECoG measures the electrical activity of the brain through electrodes embedded in a thin plastic pad that is placed on the cortex. Because ECoG is implanted inside the skull but rests outside the brain rather than within the grey matter, it has a lower risk of forming scar tissue in the brain than fully invasive BMIs such as an implant of electrode arrays. More recently, these invasive techniques have also been successfully applied to human subjects; e.g., reconstruction of hand motions from neuronal activity in the motor cortex (Collinger et al., 2013; Hochberg et al., 2012, 2006; Yanagisawa et al., 2012). Although these invasive techniques have the advantage of higher spatial resolution, better signal-to-noise ratio, and a wider frequency range, they face substantial technical difficulties and entail clinical risks (e.g., infection and damage to the brain). For the implantation of electrodes in the cortex, developing bio-compatible devices that are safe and stable over long periods of time is an ongoing challenge. Thus, compared with these invasive recordings, non-invasive methods are desirable for broad use in BMIs owing to their longevity, cost, and safety being free from surgical operations.

Most non-invasive modalities measure either electrophysiological or hemodynamic signals. Electrophysiological recordings detect changes in electric or magnetic fields induced by neural activity. The hemodynamic signal reflects neural activity, in which energy consumption is correlated with blood oxygenation. Some major candidates of these non-invasive modalities (i.e., EEG, NIRS, MEG, and fMRI) are provided below (Fig. 2.1).

EEG

EEG measures voltage fluctuations on the scalp surface, and is most widely used for BMIs in realistic environments (Birbaumer et al., 1999, 2003; Do et al., 2013; Doud et al., 2011; Hinterberger et al., 2003; Iturrate et al., 2009; Pfurtscheller et al., 2003; Rebsamen et al., 2010; Tanaka et al., 2005) because of its high portability and measurability in these environments. The origin of EEG signals is electrical dipole activities of neuronal cell assemblies in the cortex. Compared with invasive modalities, EEG recordings have a much lower spatial resolution, but the temporal resolution can be high. This is because EEG measures brain activity on the scalp through the brain, cerebral spinal fluid (CSF), and skull, and thus suffers from *volume conduction*, which represents the manner by which the source current spreads through these tissues and is determined by the geometric (i.e., surface shapes) and electrical resistivity of the tissues (Wolters et al., 2006). Therefore, EEG represents the spatially averaged source activity

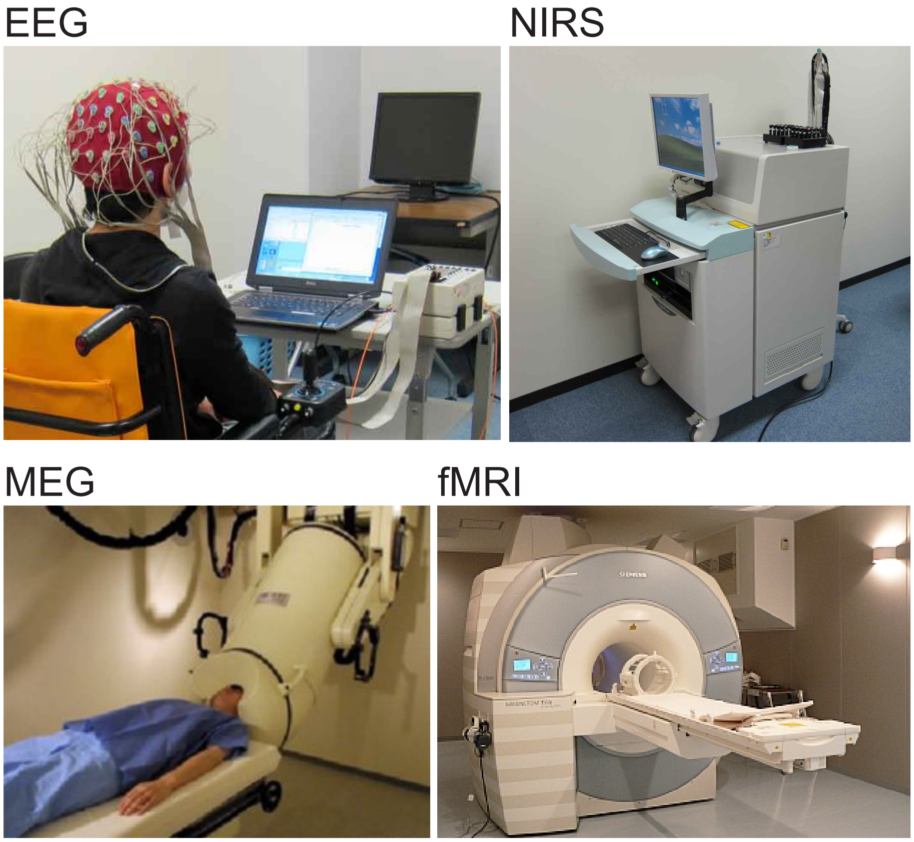


Fig. 2.1 Some major examples of non-invasive measurement modalities; i.e., electroencephalography (EEG), near-infrared spectroscopy (NIRS), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Photographs of the MEG and the fMRI were taken by the National Institute of Information and Communications Technology (NICT, *http://www.nict.go.jp/*) and Brain Activity Imaging Center, Advanced Telecommunications Research Institute International (BAIC, ATR, *http://www.baic.jp/*).

occurring within large parts of the underlying cortex, about 10–40 cm^2 of the cortical sheet, and thus each electrode reflects the source activity in tissue containing on the order of 100 million to one billion neurons (Wolpaw and Wolpaw, 2012). This is why EEG has relatively poor spatial resolution. The contamination of measurement signals originating from regions not related to the task targeted by the BMI can deteriorate the BMI's performance.

MEG

Neural activity is also reflected in magnetic field fluctuations that can be measured non-invasively by MEG (Cohen, 1968; Hämäläinen et al., 1993). MEG is most sensitive to cortical dipoles tangential to the scalp (Cuffin and Cohen, 1979). Compared with EEG, MEG has a higher spatial resolution because the magnetic fields measured with MEG are not affected by volume conduction and can resolve finer structures than EEG (Cuffin and Cohen, 1979; Grynszpan and Geselowitz, 1973). In contrast to its advantages of higher spatial resolution and successful application to BMIs (Buch et al., 2008; Chan et al., 2011; Kauhanen et al., 2006; Mellinger et al., 2007; Waldert et al., 2008), MEG technology is expensive and cumbersome. It must be used in a magnetically shielded room to avoid external magnetic signals, including the Earth's magnetic field, and the coils must be supercooled with liquid helium. These practical disadvantages make it a method that can be supportive of, but not primary to, BMI development in natural conditions.

fMRI

fMRI is the most popular modality for non-invasively measuring the hemodynamic response in the entire brain (Friston, 2009; Ogawa et al., 1990), and is most widely used in neuroscience and neurodecoding studies (Haxby et al., 2001; Horikawa et al., 2013; Kamitani and Tong, 2005; Shibata et al., 2011). fMRI measures the blood-oxygen level dependent (BOLD) signal that correlates closely with electrical activity in the cortex (Hermes et al., 2012; Ramsey et al., 2006). As neural activity increases, oxygen is carried via hemoglobin molecules in the blood (oxygenated hemoglobin, oxyHb), and then is consumed by the cells (deoxygenated hemoglobin, deoxyHb). The important aspect for neuroimaging is that oxyHb and deoxyHb have different magnetic properties. This gives rise to the BOLD signal (Ogawa et al., 1990), which can be detected by fMRI. The fMRI signal is a complex combination of blood oxygenation, blood flow, and blood volume (Buxton et al., 2004), and the relationship between neural and fMRI signals is still controversial (Logothetis, 2008). The best spatiotemporal resolution for whole-brain imaging is obtained using fMRI, and is on the order of 2 mm (Shmuel et al., 2007; Sirotin et al., 2009). The temporal resolution of fMRI is about 1 Hz, and is limited in principle by the fact that the hemodynamic response itself is slow; it is much slower than the signals measured by electrical measuring methods. An advantage of fMRI analysis is that subjects' brain scans can be co-registered with template brains, which makes multi-subject analysis possible. Although fMRI is likely to be most useful as a noninvasive method for localizing brain function in the entire brain, it is very expensive, cumbersome, and technically complex. It is also not portable. Thus fMRI cannot be applied to BMIs working in real environments.

NIRS

As with fMRI, NIRS also measures the hemodynamic response non-invasively using near-infrared light, which can travel through the intact skull (Boas et al., 2004; Jobsis, 1977; Villringer and Chance, 1997). NIRS devices consist of light-source and detector pairs positioned on the scalp. When the light source shines near-infrared light of a particular wavelength through the skull and scalp, the light passes through to the brain, is reflected back out of the brain, and is detected by the detector. As oxygenated and deoxygenated blood have different wavelength absorption characteristics, neural activity can be measured by examining the reflected light. Whereas fMRI can image the entire brain including deep subcortical structures, fNIRS is limited to the top few millimeters of the cortex, just below the skull. It has relatively lower spatial resolution (on the order of centimeters) than that of fMRI, but slightly better spatial resolution than that of EEG. Its temporal resolution is limited to several seconds primarily because of the slow response of the hemodynamic activity. Despite its limitations, NIRS has some promise as a method for BMI apprecations; fMRI for example is very expensive, cumbersome, and technically complex, whereas fNIRS is inexpensive, relatively easy to use, and portable (Villringer and Chance, 1997; Wolf et al., 2007). Thus, the optical imaging method of NIRS is a sensible alternative for measuring hemodynamic activity in BMIs in real environments (Coyle et al., 2007; Falk et al., 2011; Power et al., 2011; Sitaram et al., 2009).

2.2 Brain functions used in BMI

What type of brain function should be used in BMIs is an ongoing issue, and several approaches and brain functions have been used to output commands to external devices. Here, some examples of the most widely used modalities are provided: P300 event-related potentials, sensorimotor rhythms, and steady-state visual evoked potentials.

P300 event-related potentials

P300 is an event-related positive deflection, that occurs around the central parietal region and is elicited by the subject's decision during a rare event. The P300 is used in BMI in the context of the *Oddball Paradigm* (Donchin and Coles, 1988; Farwell and Donchin, 1988), i.e., 1)

A subject is presented with a series of events, each of which falls into one of two classes. 2) The events that fall into one of the classes are less frequent than those that fall into the other class. 3) The subject performs a task that requires classifying each event into one of the two classes. The P300 is elicited by events of the less-frequent class, and its amplitude depends on the subject's attention. Thus, we can decode the choice (decision) of the subject by detecting the P300. P300-based BMIs have been successfully used because of their simplicity, inexpensiveness, and stable performance (Birbaumer et al., 1999; Farwell and Donchin, 1988; Hoffmann et al., 2008; Kübler and Neumann, 2005; Nijboer et al., 2008; Sellers et al., 2006, 2010; Vaughan et al., 2006). However, they have relatively slow communication rates, and need external visual/auditory devices to provide subjects with stimuli for eliciting the P300.

Steady-state visual evoked potentials

Visual evoked potentials (VEPs) are one kind of EEG evoked potential (EP), and are evoked by a sudden visual stimulus. Steady-state VEPs (SSVEPs) are stable voltage oscillations elicited by rapid repetitive stimulation such as a strobe light or a pattern-reversing checker box presented on a monitor. They are induced around the primary visual cortex. In the standard SSVEP-based BMI, the user views a set of stimuli, which are placed at different locations in the visual field and flash at different rates (e.g., 8 Hz, 12 Hz, and 15 Hz). When the user looks at one of the stimuli, the frequency spectrum corresponding to that stimulus shows peaks, and can be detected from the power spectral density of EEG around the primary visual cortex. As with P300-based BMIs, SSVEP-based BMIs have also been successfully used because of their robustness (Allison et al., 2008; Brunner et al., 2010; Lalor et al., 2005; Middendorf et al., 2000; Müller-Putz et al., 2005; Pfurtscheller et al., 2010); however, they need external visual devices to show visual stimuli to the user.

Sensorimotor rhythms

Sensorimotor rhythms (SMRs) are electrical oscillations recorded over the sensorimotor cortices. SMR-based BMI is the most widely used modality for BMIs (Blankertz et al., 2007; Hema et al., 2011; Leeb et al., 2007; McFarland et al., 2010; Pfurtscheller et al., 1993, 2000; Pfurtscheller and Neuper, 2006; Wolpaw and McFarland, 2004), and is based on much evidence that the execution or imagination of different parts of the body induces changes in rhythmic activity recorded over corresponding regions of sensorimotor cortex (Holmes, 1999; Neuper et al., 2006, 2009; Pfurtscheller and Aranibar, 1979). The point-to-point correspondence between body parts and specific cortical regions is called somatotopy. The SMR-based BMIs translate the imagined particular actions (e.g., left hand movement, right hand movement, and foot movement) into outputs such as cursor movement. Compared with P300 and SSVEP-based BMIs, SMR-based BMIs have an advantage in that they do not require external devices to provide the user with visual stimuli. However, the performance of SMR-based BMI is closely related to the user's ability to imagine the motor movements.

Chapter 3

Decoding from cortical currents estimated from EEG and NIRS

3.1 Introduction

For practical BMIs, EEG and NIRS are the only current methods that are non-invasive and available in non-laboratory environments. However, the use of EEG and NIRS involves certain inherent problems. EEG signals are generally a mixture of neural activity from broad areas, some of which may not be related to the task targeted by BMI, hence impairing BMI performance. NIRS has an inherent time delay as it measures blood flow, which therefore detracts from practical real-time BMI utility.

Source current localization, which is to reconstruct cortical currents based on measurements outside the scalp, has a potential to address these issues in EEG-NIRS-based BMIs working in real environments. Compared with decoding directly from EEG or NIRS measurements, that from cortical currents has several advantages (Toda et al., 2011). First, EEG sensor signals are mapped onto currents on the cortex, each of which possesses information specific to a particular cortical region, thus avoiding the volume conduction effect. In addition, artifacts evoked by, e.g., eye movements can be eliminated by incorporating extra dipoles during the source current estimation (Morishige et al., 2009). By estimating cortical currents and separating task-relevant brain activities, improvement in BMI performance can be expected. Second, we may be able to obtain brain activities with higher spatial resolution by assuming a large number of cortical current dipoles compared to that of EEG sensors. Third, a large body of neuroscientific knowledge can be used for validating the reconstructed cortical activities and hence BMI decoders based on them. Cortical currents invoked by performing a task can be analyzed based on the function of active brain regions. Since activity in a specific brain regions would induce both cortical current and blood flow, we can compare the task-relevant cortical current sources and the fMRI-based neural correlates involved in the same task. Despite such advantages, cortical source current estimation is an ill-posed inverse problem because many different source configurations can generate the same EEG observations (Grech et al., 2008; Michel et al., 2004). Therefore, some prior assumptions are required to obtain a unique solution; e.g., L1-norm minimization method (Uutela et al., 1999), L2-norm minimization method (Wang et al., 1992), LORETTA (Pascual-Marqui et al., 1994), and the Wiener filter or Bayesian inference (Dale et al., 2000; Kajihara et al., 2004; Phillips et al., 2002; Schmidt et al., 1999) which resolved the ill-posedness by using fMRI data as prior information on the source current variance. More recently, VBMEG (Sato et al., 2004; Yoshioka et al., 2008) has been proposed to incorporate the prior information as a hierarchical prior: the blood-flow information from fMRI is substituted into the parameter specifying the probability distribution of the current variance rather than the variance itself, and the variance is estimated through the variational Bayesian estimation procedure. This constitutes placing a soft constraint on the variance, and thus is robust to the vulnerability of inaccuracies in prior information.

The simultaneous use of NIRS measurement is a promising solution for real-environment BMI as it can reduce the ill-posedness of the EEG source current estimation and estimate reasonable activation patterns. NIRS measures the concentration changes of oxygenated and deoxygenated hemoglobins (oxyHb and deoxyHb, respectively) in the superficial layers of the cortex (Villringer et al., 1993). Thus, NIRS can detect active, nonactive, and deactive cortical regions via blood flow caused by neural activities that also induce EEG signals. Aihara et al. (2012) showed that, by applying VBMEG to EEG measurement during a finger tapping task, cortical currents could be estimated better by incorporating NIRS prior than solely from EEG, though EEG and NIRS were measured by different but similar tasks.

Training a BMI classifier using relatively few data has an issue on the machine learning side. Estimated current dipoles on the cortical surface are high-dimensional in general, having more degrees of freedom as compared to the number of available data points, thus causing an overfitting problem (Hastie et al., 2009; Kriegeskorte et al., 2009). It is therefore important to reduce the dimensionality by selecting in a sparse manner current dipoles that are relevant to the task of interest. Sparse logistic regression (SLR; Yamashita et al., 2008) estimates the weight parameters of logistic regression with the automatic relevance determination (ARD) hierarchical prior via Bayesian inference, and can automatically and sparsely select informative features for decoding (BMI classification) (Miyawaki et al., 2008; Shibata et al., 2011). Thus, by using SLR for decoding after the cortical current estimation, we can effectively select the current dipoles having region-specific information related to the task.

Spatial attention is often used as a task for BMI because it is natural in real environments and justifiable from existing neuroscience/BMI studies. Because BMI users may not be able to move their eyes, covert spatial attention is a standard and natural way to decode the direction of intended control. In particular, it would be convenient to control objects in the visual domain such as a cursor on a computer screen. Recent studies have reported that subjects' covertly attended direction could be decoded from brain activities even in a single trial. For instance, Kelly et al. (2005a,b) showed that shifts in covert spatial attention between the left and right visual hemifields can be decoded from EEG's alpha-desynchronization in the posterior sites contralateral to the attended hemifield; the average accuracy for off-line decoding was 73% (Kelly et al., 2005a) and for on-line decoding was 62% (Kelly et al., 2005b).[*1] This finding has been followed by an MEG study (Van Gerven et al., 2009) with a reported off-line decoding accuracy of 61% on average, and an fMRI study Andersson et al. (2011) in real-time with an accuracy above 90%.

In this study, we propose a novel methodology for decoding covert spatial attention from cortical currents estimated from EEG with NIRS prior information with an SLR classifier. Structural MRI data are also used to provide accurate head models. This is the first study applying cortical current estimation from EEG with NIRS prior information to BMI decoding. We introduce a novel form of Bayesian prior to capture event related desynchronization (i.e., modulations in frequency power) which is a well-known phenomenon during many tasks involving higher-order functions including spatial attention. A major advantage of our new procedure is

---

[*1] This percentage was not directly reported in Kelly et al. (2005b); however we calculated their average accuracy by dividing the number of correct selections by the sum of the correct and incorrect selections (not including inconclusive trials), which are shown in Table 1 of their article.

that it can examine activation patterns on the cortex in light of existing neuroscientific knowledge. Such verification would also imply the possibility that the EEG-NIRS-based decoding technique contributes to studies of brain functions in non-laboratory and natural conditions.

3.2 Method

3.2.1 Subjects

Experiments were conducted on eight right-handed males between 20 to 40 years of age (mean: 24.6, SD: 6.4) who had normal or corrected-to-normal vision. All subjects gave written informed consent for the experimental procedures, which had been approved by the ATR Human Subject Review Committee.

3.2.2 Experimental setup

The subjects were seated in a comfortable chair 1 m away from a 19-inch display for visual presentation. A white cross was displayed at the center of the display as a fixation point. Visual stimuli of left and right bars were centered at an eccentricity of 8°. The size of the bars was 1.3°. The head position of the subject was fixed by chin and forehead supports. A keyboard for response was placed in front of the subject. This experiment was conducted in a dark room.

3.2.3 Task for subjects

We used a spatial attention task in which subjects attend to the left or right following instruction by the visual stimulus. This task was modified from Tootell et al. (1998). Figure 3.1A shows the time course of epochs in a block. One block consisted of two epochs: Attention (8 s) and Control (4 s). One experiment consisted of 8 sessions, and each session consisted of 24 blocks.

In the Attention epoch, two white flashing bars were presented repeatedly in a rapid stream until the end of the epoch. These flashing bars were presented for 100 ms followed by an inter-stimulus interval (600–800 ms) where no bar was presented. At each appearance, the orientations of each bar were selected randomly with equal probability from the following: -30°, 0°, and 30°. To inform the subjects of the target direction of an epoch, one bar was colored red and the other was colored green at initial presentation in the epoch. The target direction (Left or Right) was pseudo-randomly selected for each Attention epoch. We informed the subjects of the target color (red or green) for each session before starting the session, and the color was fixed during the session. To preclude issues from the difference in target color, the order of color was pseudo-randomly selected for each session. During the Attention epoch, subjects were instructed to fixate their eyes to the fixation point, and covertly and continuously orient their attention to a single bar that was cued by a target color at the initial presentation of the bars. To check whether the subjects continuously attended to the cued direction, we asked the subjects to only press the button immediately after the target bar was vertical (0°). Flash timings and bars' orientations were designed to be difficult for the subjects to respond without the attention to the target direction.

The Control epoch always precedes the Attention epoch. Visual stimuli during the Control epoch are the same as those during the Attention epoch, i.e., repetitive presentations of left and right bars. The bars presented for the first time in each Control epoch were colored blue. The Control epoch was inserted to cancel out the difference of visual stimuli before and after the onset of the Attention epoch, and to focus only on the modulation caused by selective spatial attention. During the Control epoch, subjects were instructed to fixate their eyes on the fixation point and to distribute their attention continuously and evenly over the two bars.

Each session consisted of 24 blocks (12 lefts and 12 rights), with short Rest epochs in between (6–10 s). In the Rest epoch, nothing was displayed on the screen. The onset of each block was cued by a short beep. During each block, subjects were instructed to steadily maintain fixation on the fixation point and to refrain from blinking as much as possible. We conducted 8 sessions for each subject with a brief break between each session. Thus, the total number of trials (Attention epochs) was 192 for each subject.

3.2.4 Data acquisition

We simultaneously recorded EEG and NIRS signals (Fig. 3.2). Using a Biosemi ActiveTwo system, the EEG recording was performed at a sampling rate of 256 Hz with a 64-electrode cap, referenced to the common mode sense (CMS) active electrode. EEG channels were placed to cover the whole head (Fig. 3.1B, left). We also measured electrooculography (EOG) to ensure that subjects did not shift their eye gaze towards the attended directions by analyzing the performances of decoders with EOG only.

NIRS recording was performed via FOIRE-3000 (Shimadzu Co., Japan), an optical imaging system. We used 15 emitters and 15 detectors, placed to cover the parietal and occipital lobes, resulting in 49 channels (Fig. 3.1B, right). An emitter irradiates near-infrared beams (wavelengths 780, 805, and 830 nm), and diffused beams sampled by a detector at 4 Hz were used to calculate changes in the concentrations of oxyHb and deoxyHb. The distance between one emitter and one detector was set at about 3 cm for each pair.

In order to construct forward models used in VBMEG, structural MRI data were recorded via a 3T Magnetom Trio MRI scanner (Siemens, Germany). T1-weighted images were acquired with the following parameters: repetition time of 2.25 s, time of echo 3.06 ms, flip angle 9°, slice thickness 1 mm, field of view of 256 $\times$ 256 mm, imaging matrix of 256 $\times$ 256, and 208 slices.

Before starting a series of experiments, the face surface and the locations of EEG sensors and NIRS probes were measured by FastSCAN (Polhemus, USA), a laser scanner and stylus. The positions of EEG electrodes and NIRS probes were coregistered to the coordinates of the T1-weighted image using the facial scan data and three reference points: the nasion, left pre-auricular point (LPA), and right pre-auricular point (RPA).

3.2.5 EEG data preprocessing

EEG and EOG signals were passed through a band-pass filter (0.5–40 Hz). The filtered data were then divided into trials. The baseline for each trial was corrected by subtracting the mean value from that obtained 2 to 0 s before the Attention epoch.

We performed trial rejection based on the subjects' task performance in each trial. We rejected trials that were not significant with $\alpha = 0.05$ under the null hypothesis that the number of correct responses was in the chance level. Since a successful subject must not push the button after non-vertical bar presentations, which occurs with 2/3 chance, the significance for the chance level is calculated based on the cumulative function of a binomial distribution with success probability 2/3. For each subject, 5 to 36 trials were rejected (subject-wise mean: 16.3, SD: 12.2). Although we did not reject trials based on EOG (because the attentional duration was too long to keep the eyes open), we removed EOG artifacts from EEG using extra dipoles, as mentioned in Section 3.2.7.

3.2.6 NIRS data analysis

Similar to the study by Aihara et al. (2012), we used the oxyHb concentration for this analysis after determining that oxyHb produced a decoding accuracy higher than totalHb with our technique.*2 For each channel, the baseline trend was removed by DCT-detrending (cutoff: 80 s), which is a standard detrending technique in SPM (Penny et al., 2006). The data were then normalized by using the standard deviation during the Control epoch. We removed the first principal component of data, which is considered to be the skin blood flow artifact (Virtanen et al., 2009). After applying these filters, the skin-level NIRS signals were projected onto the cortical surface using structural MRI and the three-dimensional positions of the probes, and interpolated using the distance along the cortical surface.*3 We then calculated t-values for differences of the NIRS activities between the Attention epochs and the Control epochs (i.e., mean(Attention epoch)$-$mean(Control epoch)$/$standard error) by using the NIRS-SPM toolbox (Ye et al., 2009) based on the general linear model with a design matrix calculated from a task schedule. Each t-value was normalized to have a maximum absolute value of 1 when substituted into the hierarchical prior of VBMEG. When visualized on the cortical surface, the (unnormalized) t-statistics were interpolated using a spatial smoothing matrix provided by VBMEG.

3.2.7 Cortical current estimation using VBMEG

VBMEG (Sato et al., 2004) is an effective method that can address the ill-posed problem of cortical current estimation from EEG/MEG sensor signals. It can yield high spatiotemporal resolution by incorporating anatomical (MRI) and functional (fMRI/NIRS) information, and has been used successfully in previous studies (Shibata et al., 2008; Toda et al., 2011; Yoshimura

---

*2 Some researchers have reported that oxyHb is more sensitive to cerebral blood flow (CBF) (Hoshi et al., 2001; Strangman et al., 2002); however, others have found that totalHb is more correlated to CBF (Huppert et al., 2006).

*3 A more precise mapping of NIRS measurements on the cortical surface may be obtained by solving an inverse problem called diffusion optimal tomography (DOT) (Shimokawa et al., 2012), but this will require denser configurations of NIRS probes.

et al., 2012; Yoshioka et al., 2008). In VBMEG, anatomical information is used to determine the locations and orientations of current sources (current dipoles), while functional information is used to provide topographical information of amplitudes of the current sources.

Using FreeSurfer (Dale et al., 2000), a polygon cerebral cortex model was constructed from the T1 structural image of the subject. The number of vertices (corresponding to current positions or dipoles) for each subject was downsampled to 1004 to avoid the overfitting problem and to reduce the computation cost of cortical currents estimation, feature extraction, and classification. Although the number of vertices in this study (1004) is relatively smaller than those used in Toda et al. (2011) ($\sim$2500 dipoles) and Yoshimura et al. (2012) ($\sim$2800 dipoles), the task was different. In reconstructing motor movement in these studies, higher densities of vertices were effective because cortical currents around the motor cortex have information on a microscopic region-specific scale (somatotopic organization). In contrast, the present study focuses on the classification of the subjects' spatially attended direction (left or right) rather than the reconstruction of motor movement. Selective spatial attention causes the broad distribution of band-power modulation (Siegel et al., 2008), thus high resolution as the motor-related activities was not needed. A single current dipole was assumed at each vertex point perpendicular to the cortical surface. For the calculation of the lead field, which describes the relationship between cortical currents and EEG measurement, the brain structure was approximated by a three-layer model composed of CSF, skull, and scalp. Obtained by FreeSurfer, the boundaries between the layers were slightly modified by using gray/white/CSF segmentation via SPM2 and morphological operations. Using the obtained boundaries, the lead fields were calculated by the boundary element method (Mosher et al., 1999; Tissari and Rahola, 2003) implemented by Sato et al. (2004).

VBMEG incorporates NIRS/fMRI information as a parameter for the hierarchical distribution of the current variance. The cortical current at each source location is considered as an unknown variable and is estimated from the EEG signal by introducing the ARD hierarchical prior using NIRS/fMRI. The ARD prior distribution is controlled by a prior mean variance parameter $\bar{\nu}_{0i}$, where i is the index of current dipoles (see Appendix A). For large $\bar{\nu}_0$, estimated currents can be large. In contrast, for small $\bar{\nu}_0$, estimated currents tend to be small. In previous studies (Aihara et al., 2012; Toda et al., 2011; Yoshimura et al., 2012; Yoshioka et al., 2008), NIRS/fMRI information is imposed on the prior mean variance parameter by the following:

$$\bar{\nu}_{0i} = \nu_{0\text{base}} \left[1 + (m_0 - 1)t_i^2\right], \tag{3.1}$$

where m_0 (≥ 1) is the magnification parameter to be tuned (Yoshioka et al., 2008), $\nu_{0\text{base}}$ is the baseline of the current variance estimated from the EEG data, and t_i is the NIRS t-value normalized to [-1, 1]. More specifically, under the assumption that all dipoles have the same prior current variance (Toda et al., 2011; Yoshioka et al., 2008), the baseline variance $\nu_{0\text{base}}$ is computed from the EEG data in the baseline interval (2–0 s before the Attention onset) by a Bayesian minimum norm method. The expression (3.1) describes that NIRS-activated regions have larger current variances compared with non-activated regions. Since Eq. (3.1) considers the squared t-values, it can only represent an increase of neural activity, which is often observed in event related potential (ERP). Indeed, previous applications of VBMEG so far have dealt with conditions where the increase in current variances is dominant: retinotopic activities evoked by visual response (Yoshioka et al., 2008) or motor activities induced by sensorimotor tasks (Aihara et al., 2012; Toda et al., 2011; Yoshimura et al., 2012).

Existing fMRI studies have shown that spatial attention causes not only activation in the parietal cortex (Corbetta et al., 2000; Hopfinger et al., 2000; Kastner and Ungerleider, 2000; Serences and Yantis, 2006) but also deactivation in the visual area (Sylvester et al., 2007; Tootell et al., 1998). This means that the convention for the prior current variance mean, Eq. (3.1), is not sufficient because it cannot incorporate deactivation; negative values of t_i are mapped to positive by the square function (t_i^2). In this study, we propose a new expression of the prior variance parameter $\bar{\nu}_{0i}$:

$$\bar{\nu}_{0i} = \nu_{0\text{base}} \cdot m_0^{t_i}, \tag{3.2}$$

where $\nu_{0\text{base}}$ and t_i are the same as those in Eq. (3.1), but m_0 is another magnification parameter different from that in Eq. (3.1). That is, $\bar{\nu}_{0i}$ is scaled by multiplying or dividing by $m_0^{|t_i|}$ according to the positivity or negativity of t_i. Thus, NIRS-activated regions ($t_i > 0$) have larger current variances compared with baseline. In contrast, NIRS-deactivated regions ($t_i < 0$) have smaller current variances compared with baseline. In 3.3.5, we compare the spatial patterns of decoder weights trained from cortical currents with Eq. (3.2) and those with Eq. (3.1). We used this proposed prior variance in the remaining analyses.

Since we did not have access to the directional knowledge of left or right when we estimated cortical currents, we used the same NIRS prior both for the left and right conditions. In other words, we used the activities of Attention $-$ Control as prior information for both conditions and did not consider the difference in NIRS activities between the left and right conditions. In the case of spatial attention task, incorporating the directional knowledge might not result in the improvement in cortical current estimation. Indeed, an fMRI study using a spatial attention task reported that brain activities for the left and right conditions relative to the control condition are similar to each other (Hopfinger et al., 2000).

VBMEG has two hyperparameters that have to be manually tuned: m_0 and γ_0 (see Appendix A). We already know that m_0 is a prior magnification parameter. The second hyperparameter γ_0 reflects the reliability (confidence) of the prior knowledge. For very small γ_0, the probability distribution of the current variances (Eq. (A.4) in Appendix A) spreads uniformly (becoming a non-informative prior) and the prior information $\bar{\nu}_{0i}$ has little influence on the current estimation. When γ_0 is large, since the distribution is concentrated around prior mean $\bar{\nu}_{0i}$, prior mean $\bar{\nu}_{0i}$ constrains the current estimation more strongly. We evaluated BMI performances for the following hyperparameters (range): $m_0 \in \{10, 100\}$ and $\gamma_0 \in \{1, 10, 100, 500, 1000\}$ ($2 \times 5 = 10$ pairs). These values are within the proper range as analyzed theoretically (Yoshioka et al., 2008), and experimentally in a visual task (Yoshioka et al., 2008) and a motor task (Aihara et al., 2012). Details of this evaluation method are described in Section 3.2.11.

Once the variance of each cortical current is estimated, we can calculate an inverse filter and estimate the cortical currents by multiplying sensor signals by the inverse filter (Sato et al., 2004). To deal with the nonstationarity of brain activities, inverse filters were first estimated for each sliding window (500 ms length with 250 ms overlap) from 2 s before to 9 s after the Attend onset under the assumption that the current variance was temporally invariant within each time window. Then two inverse filters were averaged in the overlapping time intervals. The data from 2 s to 0 s before the Attend onset were used to estimate the baseline of the current variance, $\nu_{0\text{base}}$ in Eq. (3.2).

In order to eliminate artifacts evoked by eye movements, we incorporated artifact dipoles ("extra dipoles") in the source current estimation problem (Morishige et al., 2009). Artifact dipoles were placed at the center of the left and right eyeballs whose coordinates were determined from the structural MRI image. Currents estimated at the extra dipoles can be projected back to the EEG signal domain by multiplying corresponding columns of the lead field matrix.

3.2.8 Cortical current estimation using VBMEG without an NIRS prior and the minimum L2-norm method

In addition to comparing the proposed decoding method with the method decoding solely from EEG sensor signals, to investigate the effects of using NIRS prior information and that of using VBMEG as a cortical current estimation method, we compared the following two methods: Hierarchical Bayesian estimation without NIRS prior and minimum L2-norm estimation (Wang et al., 1992). Thus, as shown in Fig. 3.1C, the former method eliminate NIRS pathway in the diagram, and the latter method replace VBMEG with minimum L2-norm estimation method.

VBMEG without NIRS prior information can be obtained by simply setting $m_0 = 1$; this will cancel the effects from NIRS t-values (see Eq. (3.2)).

Minimum L2-norm estimation is a conventional cortical current estimation method that resolves the ill-posedness of the inverse problem by regularizing solutions with their L2-norm. It does not estimate dipole-wise variances; hence it produces less-sparse cortical currents compared to VBMEG. The minimum L2-norm method can be obtained by constraining the variances of all the dipoles to be the same value, which was estimated from the Attention period.

3.2.9 Feature extraction

We used alpha (8–14 Hz) and beta (14–30 Hz) bands as the feature values for the classifier; the use of these frequency bands can be justified by prior findings on spatial attention (Kelly et al., 2005b; Siegel et al., 2008; Van Gerven et al., 2009).[*4] The band powers from 0 s to 8 s after the Attend onset were extracted from each channel (EEG) or vertex (current) using the multitaper power spectral estimation and transformed to log-band powers. Existing reports (Kelly et al., 2005b; Van Gerven et al., 2009) suggested that spatial attention induces contralateral alpha-desynchronization around the visual area, and could be decoded from alpha-band modulations around occipito-parietal regions. In addition, Siegel et al. (2008) used MEG to show that spatial

---

[*4] The optimization of frequency bands for each subject may improve accuracy (Ang et al., 2008); however, we used the same bands for all the subjects.

attention also modulates beta bands around occipito-parietal regions.

3.2.10 Classification

We used SLR (Yamashita et al., 2008), which estimates the weight parameters of logistic regression with the ARD hierarchical prior, and automatically selects a small number of relevant features.

3.2.11 Evaluation of classifier

We used two criteria for evaluating classifiers, namely, the cross-validation (CV) accuracy and the nested CV (NCV) accuracy. CV was used for evaluating the decoding performance of a classifier with given hyperparameter values, whereas NCV was used for evaluating the generalization performance of a classifier without bias in hyperparameter selection. When we decoded solely from EEG or from EOG, we only applied CV because there were no hyperparameters.

For the CV procedure, seven sessions out of eight were used for training, and one session was used for testing. Using different partitions, we repeated the training and testing with given hyperparameter values eight times, and then averaged the eight testing accuracies to produce a CV accuracy for each subject. The mean of CV accuracies across subjects defines *the CV decoding accuracy* for the classifier. Since the CV decoding accuracy depends on the hyperparameters, this procedure for calculating the CV decoding accuracy was repeated for each of the 10 pairs of hyperparameters.

For the NCV procedure, the same eight-fold partition of the data for CV was applied; however, a two-stage procedure consisting of an outer CV and inner CV was employed. The outer-CV stage works much like the standard CV procedure, but in its training phase it selects optimal hyperparameters as opposed to using given values. To select the best hyperparameters from the 10 settings, the inner-CV procedure partitions the outer-CV training data consisting of seven sessions into (1) inner-CV training data consisting of six sessions, and (2) inner-CV testing data consisting of the other session. The inner-CV accuracy is the average of the seven testing accuracies. We selected the hyperparameter values that achieved the maximum inner-CV accuracy and used them for obtaining a classifier that was re-trained using the entire outer-CV training data, and the classifier was used for evaluating outer-CV test accuracy. There is no single “optimal” hyperparameter setting for NCV. *The NCV decoding accuracy* for the classifier is the mean of the subject-wise NCV accuracies.

Figure 3.1C shows the concepts contrasting the difference between the proposed and the EEG only (‘sole-EEG’) method. The proposed method was evaluated based on its decoding accuracy obtained by using the cortical currents estimated from the EEG signals with the NIRS prior. The sole-EEG method was evaluated by its accuracy of decoding only from the EEG signals. The above methods, as well as VBMEG without NIRS prior and the minimum L2-norm method, all used SLR as a classifier.

We examined which anatomical regions contributed to the classification of spatial attention by analyzing trained weights from SLR for EEG sensors and current sources. For visualization, the weights were averaged across the eight classifiers. The sensor weights were interpolated on the skin surface, and the source weights were spatially smoothed on the cortical surface.

3.3 Results

3.3.1 Evaluation of decoding accuracy

The NCV decoding accuracy of attended directions was compared between the two cases: (a) decoding from the cortical currents estimated with VBMEG, (b) decoding solely from the EEG sensor signals. We found that the NCV accuracy in case (a), the proposed method with VBMEG and SLR, was significantly higher than that in case (b) (79.1% vs. 71.4%, $p = 0.030$ with two-sided paired t-test, Fig. 3.3).

In addition to the main result of comparing between (a) and (b), the proposed method (a) was compared with the following two cases: decoding from the cortical currents estimated with (c) VBMEG without NIRS prior, and (d) minimum L2-norm. Without using NIRS prior estimation, case (c), the NCV decoding accuracy (77.3%) was higher than that by the sole-EEG method (77.3% vs. 71.4%, $p = 0.069$ with two-sided paired t-test) but was significantly decreased compared to the proposed method (77.3% vs. 79.1%, $p = 0.046$ with two-sided paired t-test). The NCV accuracy obtained using the minimum L2-norm method, case (d), was 75.9%, which was higher than the accuracy of the sole-EEG method (75.9% vs. 71.4%,

$p = 0.16$ with two-sided paired t-test) but was significantly worse than that of VBMEG case (a) (75.9% vs. 79.1%, $p = 0.011$ with two-sided paired t-test).

In Fig. 3.3, accuracy for EEG was obtained without reducing artifact from EEG signals; however, simple artifact reduction did not result in a significant improvement. We tested the decoding accuracy from EEG signals whose artifactual component was attenuated by subtracting signals induced from the extra dipoles, and found no significant improvement from the sole-EEG method (73.7% vs. 71.4%, $p = 0.16$ with two-sided paired t-test).

The subjects' spatially attended direction could not be decoded from the EOG signals. The CV decoding accuracy from the EOG signals was 51.5% (SD: 5.7). The subjects did not achieve decoding accuracies significantly above the chance level ($p > 0.01$, one-sided paired t-test). Thus, we conclude that the EOG artifact did not have enough information for decoding spatial attention.

We used NCV to evaluate the statistical significance of the difference in the decoding accuracy between the proposed method and the sole-EEG method in this subsection. In the following results, we used a fixed hyperparameter pair for each subject, which maximized the CV accuracy of the subject.[*5]

3.3.2 EEG sensor signals and NIRS activities

Figure 3.4 shows the estimated cortical currents and EEG and NIRS activities of S2 who showed the most interpretable effects of the proposed method. The patterns of the estimated cortical currents (Fig. 3.4C) showed relatively obvious hemispheric lateralization of alpha- and beta-band powers compared with EEG signals (Fig. 3.4A).

The t-values (Attend right − Attend left) of extracted alpha- and beta-band log powers of the EEG sensor signals are shown in Fig. 3.4A. We did not observe hemispheric lateralization (alpha-desynchronization in the posterior sites contralateral to the attended direction), which has been previously demonstrated in a covert spatial attention task (Thut et al., 2006; Worden et al., 2000; Wyart and Baudry, 2008). The representative subject shows negative values (i.e.

[*5] Although this hyperparameter selection would have introduced a bias to the decoding accuracy, we used that hyperparameter not for evaluating the decoding accuracy but for evaluating the spatial patterns of activities and weights of the classifiers.

relative suppression of the alpha-band power in the right condition or relative enhancement of the alpha-band power in the left condition) not only in the left hemisphere but also in the right hemisphere. As a result, we did not see hemisphere-specific modulation of the alpha band around the posterior sites directly from the EEG signals. We also did not see hemispheric lateralization around the posterior sites in the beta band. As suggested by Siegel et al. (2008), visual stimulus to a hemifield (left or right) that is spatially attended induces beta-synchronization in the posterior sites contralateral to the attended hemifield. Although the subject here showed slight contralateral beta-synchronization around the parietal area, beta-synchronization was not evident around the occipital area.

The NIRS activities were reasonable in the light of existing studies of spatial attention with fMRI (Fig. 3.4B). This figure shows t-values (Attention − Control) of the NIRS activities from the same subject. We can see the activation around the parietal region, which is consistent with related studies (Corbetta et al., 2000; Hopfinger et al., 2000; Kastner and Ungerleider, 2000; Serences and Yantis, 2006). In addition, we can see deactivation around the visual area, which is also consistent with prior findings (Sylvester et al., 2007; Tootell et al., 1998). This spatial pattern of activation and deactivation was consistent even in a group analysis (data not shown). These results indicate the importance of considering both positive and negative values of NIRS activity for cortical current estimation.

3.3.3 Estimated cortical currents using VBMEG

The patterns of the estimated cortical currents of the same subject showed relatively obvious hemispheric lateralization of alpha- and beta-band powers compared with EEG signals (Fig. 3.4C). Similar to Fig. 3.4A, Fig. 3.4C shows the t-values (Attend right − Attend left; $p < 0.05$; uncorrected) of the extracted alpha- and beta-band log-powers on each vertex of the estimated cortical currents. Unlike the EEG analysis, the negative values around the right parietal region did not emerge in the alpha band, and there were clearer contralateral alpha-desynchronization patterns around the left visual and parietal cortex. In addition, this subject showed alpha-band activities around the frontal eye field (FEF). Indeed, FEF is known to be involved in the control of spatial attention through the modulation of alpha-band activities (Capotosto et al., 2009). Thus, these activities in FEF are thought to be related to selective spatial attention. Moreover, the beta-band activities showed synchronization in the posterior sites contralateral to the attended hemifield, which is consistent with Siegel et al. (2008). These results suggest that we have achieved appropriate isolation of the cortical currents from the EEG signals avoiding the contamination of EEG signals.

3.3.4 Anatomical localization of source currents contributing to decoding

We plotted the weights of the classifiers to analyze the anatomical localization of the source currents that contributed to decoding (Fig. 3.5). We found that the weights of the classifiers were concentrated on regions consistent with neuroscience knowledge (Fig. 3.5, left).

The weights for the cortical currents classifier were located in areas justifiable from knowledge of neuroscience (Fig. 3.5, center). In the alpha band, the weights were mainly concentrated near the left intraparietal sulcus (IPS) and visual cortex. This spatial pattern was consistent with existing findings about spatial attention, which suggested that the IPS is important for selective spatial attention (Corbetta et al., 2000; Hopfinger et al., 2000). In the beta band, weight values were located around the left occipital cortex and the parietal cortex, which are slightly far from the IPS. These weights had opposite signs compared with those for the alpha band. This is consistent with the opposite activities between the alpha and beta bands (Fig. 3.4C). Our results imply that the decoding accuracy of this subject was improved by isolating task-relevant brain activities, and by decoding with an appropriate selection of cortical currents related to spatial attention.

These tendencies were consistent even at the group analysis (Fig. 3.5, left). We can see that the weight values of the alpha band were mainly located around the left and right IPS with opposite signs. As mentioned above, IPS is thought to be related to selective spatial attention, and cause contralateral alpha-desynchronization. These results indicate that cortical currents were estimated appropriately, and the classifiers could select informative cortical currents, i.e. IPS, even at the group level.

Weight values of the EEG classifier for the same subject were distributed over the scalp (Fig. 3.5, right). Large positive weights for the beta band were observed around the parietal area (P2), and large negative weights were located around the frontal area (AFz). Additionally, the weight values were spread over a wide area ranging from occipital to frontal, and from alpha to beta bands. Thus, it is difficult to interpret the specific brain regions important for decoding of spatial attention.

3.3.5 Comparison of proposed variance prior, conventional variance prior, and no variance prior

We contrast the three prior variance models by comparing spatial patterns of the classifier weights. Figure 3.6 shows the spatial patterns of the weights of classifiers trained using cortical currents estimated with the proposed prior variance model, Eq. (3.2), the conventional prior variance model, Eq. (3.1), and the no prior variance model obtained by substituting m_0 of Eq. (3.1) by 1. From this figure, we can see that the spatial distribution of the classifier's weight with the proposed model is more concentrated in specific regions around IPS compared to those by the conventional method and the no prior method. In other words, the cortical distribution of the classifier weights were more scattered with the conventional model, and farther from IPS with the no prior model than the proposed model (Fig. 3.6). This signifies that the use of NIRS prior information and the new relationship between NIRS and the hierarchical prior can result in focal estimation consistent with neuroscience knowledge.

3.4 Discussion

In this study, we decoded spatially attended directions (left or right) of human subjects from source currents estimated on the cortical surface from EEG, NIRS, and structural MRI data, and found that the decoding accuracy was superior compared to that solely decoded from sensor signals, and moreover, the activation patterns on the cortex were reasonable. We employed VBMEG (Sato et al., 2004; Yoshioka et al., 2008) to estimate cortical currents from EEG signals with NIRS prior and SLR (Yamashita et al., 2008) to decode spatially attended directions with automatically selecting relevant cortical currents.

We posit that the selective use of informative cortical currents relevant to the processing of spatial attention resulted in superior decoding accuracy. First, cortical currents estimated by solving the inverse problem using VBMEG successfully isolated task-relevant brain activities from non-related ones (Fig. 3.4C), as contamination in EEG signals caused by the volume conduction effect was eliminated (Fig. 3.4A). Furthermore, assuming extra-current dipoles outside the brain (on the eyeballs) allowed for the separation of EOG artifacts. The effectiveness of VBMEG in mapping raw measurements into neurologically meaningful quantities has been previously demonstrated: Yoshioka et al. (2008) successfully reconstructed retinotopic activities in the primary visual areas from MEG signals with the fMRI prior, and Aihara et al. (2012) estimated cortical currents relevant to motor activities from EEG signals with the NIRS prior. The results (Fig. 3.4) indicated that although EEG signals were highly contaminated by the volume conduction effect (and thus decoding directly from the EEG signal would be difficult), the spatial patterns of alpha- and beta-band activities of estimated cortical currents showed clear hemisphere-specific activities (Fig. 3.4C) compared with those of EEG signals (Fig. 3.4A). Although this subject showed contralateral beta-synchronization around the parietal area, some other subjects showed contralateral beta-desynchronization around the parietal area. One possible reason for this inconsistency between the subjects may be because of the relatively lower range of the beta band (14–30 Hz) that was used for analysis. This lower range of the beta band is relatively near the alpha band compared with a previous report (15–35 Hz; Siegel et al., 2008). Although the subject S2, who was a typical instance, showed contralateral alpha-desynchronization only around the left visual and parietal cortex, some other subjects showed that around both left and right regions.

The second possible factor for the performance improvement is the selective use of features using SLR (Fig. 3.5, left). From cortical currents, which have a dimensionality (1004) over 10 times higher than the EEG sensors (64), SLR selected sparse features. The spatial distribution of the estimated weights of SLR was focal and had modes around IPS (Fig. 3.5, center). Indeed, the contribution of IPS to the functionality of selective spatial attention has been shown in several studies (Corbetta et al., 2000; Hopfinger et al., 2000). This suggests that the classification was performed based on the modulations of neural activities caused by selective spatial attention. Even at the group level (Fig. 3.5, left), the main positive and negative weights for the alpha band were located around the IPS, and other 'non-main' weights were observed in a few areas including the parietal and occipital cortices. This means that SLR prevented the overfitting problem and decoded the subjects' spatial attention by reasonably selecting the cortical currents relevant to the spatial attention. In contrast, when spatial attention was predicted solely from the EEG sensor signals, the weight values of the classifier were scattered over a wide range of areas ranging from occipital to frontal, and of frequency bands from alpha to beta (Fig. 3.5, right), and thus it was difficult to infer which regions were important for classification. These results showed the effectiveness of combining VBMEG and SLR for decoding from NIRS-EEG.

By comparing our model with VBMEG estimation without NIRS prior and the minimum L2-norm method, we found that both the NIRS prior information and the sparseness induced by VBMEG were effective in improving the decoding accuracy of spatial attention. In addition, we found that the classifier failed to place its weights on IPS without NIRS prior information (Fig. 3.6, right). These results indicate that using NIRS-prior can enhance the scientific validity of decoders that are based on cortical currents.

It is clear that both the positive and negative activities of fMRI/NIRS in the hierarchical prior distribution are necessary, because negative t-values were observed around the occipital area in our task (Fig. 3.4B). Thus, Eq. (3.2) is preferred to the existing standard in the VBMEG literature [i.e., Eq. (3.1)]. The NIRS t-values shown in Fig. 3.4B have both positive and negative peaks, and the negative peaks have magnitudes similar to the positive peaks; these facts support the validity of the novel model. This mixture of positive and negative activities can be the case for many tasks involving higher-order functions; e.g., motor movement (Nirkko et al., 2001), spatial attention (Sylvester et al., 2007; Tootell et al., 1998) and working memory (Tomasi et al., 2006). Although the NCV accuracy with the novel model exceeded that with the model without NIRS prior (79.1% vs. 77.3%, $p = 0.046$ with two-sided paired t-test, cases (a) and (c) in Section 3.3.1, respectively), the NCV accuracy with the conventional model was not significantly better than that with the method without the NIRS prior (78.6% vs. 77.3%, $p = 0.068$ with two-sided paired t-test). This result supports the effectiveness of the proposed model. In-depth comparisons between the proposed prior variance model, Eq. (3.2), and the conventional prior variance model, Eq. (3.1) are provided in 3.3.5.

The relationship between the number of current dipoles and decoding accuracy needs further investigation. In this study, we used the number of current dipoles (1004) in the cortical current estimation stage. Although this is considerably larger than that of EEG sensors, the result did not suffer from an overfitting problem. It is worth mentioning that the subjects' attended directions were decoded with significantly higher accuracy than that from EEG sensor signals alone (79.1% vs. 71.4%, $p = 0.030$ with two-sided paired t-test); however, the proposed method did not always outperform the sole-EEG baseline (Fig. 3.3A). Yoshimura et al. (2012) reported that the accuracies for decoding flexor and extensor muscle activities with for all of their subjects significantly improved with $\sim$2800 dipoles, an analysis (not included) of our data showed that there was no performance gain for the spatial decoding task even after increasing spatial resolution. Based on these results, an appropriate number of current dipoles might depend on the modality to be decoded. Cortical current estimation may be improved by using a head model with more detailed structure; e.g., inclusion of gray and white matters in addition to CSF, skull, and scalp. According to Güllmar et al. (2010), taking into account the anisotropy of white matter may improve the localization accuracy up to about 5 mm. With our settings, the mean distance between nearest dipoles is 7.3 mm and thus a denser dipole configuration is required to receive the benefits from a detailed head model. The relationship between the marginal quality of the cortical estimation and the decoding accuracy of spatial attention is still unclear and needs further investigation.

Our study can be extended to real-time decoding in real environments. Although we measured EEG and NIRS simultaneously, NIRS information was only used as prior knowledge to analyze EEG signals to produce cortical currents. In other words, we have to process EEG data in real time but it is sufficient to have off-line NIRS data (used for calculating the prior mean variance parameter). Fazli et al. (2012) simultaneously measured EEG and NIRS in addition to multi-modal decoding but without source localization. The authors decoded two classes of motor movement and motor imagery (left hand or right hand) by using a meta-classifier, which combined the outputs of three classifiers trained separately for EEG and NIRS (oxyHb and deoxyHb), and achieved significant improvement of classification accuracy than that with sole-EEG. The main difference between our present study and Fazli et al. (2012) is that their meta-classifier needed immediate output from the NIRS classifier, and therefore they would suffer from the time delay of NIRS activities after a task onset. Other studies that use a similar methodology have the same limitation (Ma et al., 2012). Since we used NIRS only as prior information, we could decode in the timescale of EEG. Moreover, when applied to real environments, the subject does not have to take on NIRS; this relaxes the measurement conditions and is beneficial both for the experimenter, in terms of time for putting on NIRS probes, and for the subject, who is released from the load of NIRS. For a reference, we applied the method by Fazli et al. (2012) to our spatial attention data and found its NCV accuracy was 69.3% with deoxyHb NIRS, which was comparable to the sole-EEG decoding accuracy of 71.4%. Some issues to bring our methodology to real environments include the reduction of artifacts both in EEG and NIRS and the acceleration of computation time.

Although NIRS measures only superficial layers, it may give sufficient information for reconstructing and decoding brain activities related to spatial attention. Indeed, the most important region for spatial attention was the IPS (Fig. 3.5, left) which is located in a relatively superficial layer.

Many EEG-based decoding methods are based on spatial filtering (Wolpaw and Wolpaw, 2012); thus, we show a comparison of the proposed method to common spatial patterns (CSPs) (Blankertz et al., 2008), a major spatial filter. We followed the settings recommended in Blankertz et al. (2008): EEG signals were band-pass filtered in the 8–30 Hz frequency band, reduced to 6 dimensions using the corresponding CSP filters, and then the log-variances of the 6-dimensional data were used as features for linear discriminant analysis (LDA). The CV accuracy of the CSP-based classifier was 72.8%, which was not significantly different compared to the sole-EEG method ($p = 0.74$, two-sided paired t-test). Spatial filtering is effectively extracting information from EEG signals but is not directly capable of localizing relevant brain activities on the cortical domain.

3.5 Summary

In this study, we proposed a novel methodology for decoding subjects' covertly attended direction from cortical currents estimated from EEG with NIRS prior information, which is applicable to real-time BMIs working in real environments. When estimating cortical currents from EEG measurement, we proposed a new NIRS-based prior variance model suitable for treating event related desynchronization caused by higher-order functions including spatial attention. In the estimation procedure, structural MRI data were also used as a head model. After the cortical current estimation, the subjects' attended direction was decoded by a Bayesian regression method whose hierarchical prior setting enabled automatic selection of informative cortical currents. Not only our new procedure outperformed major previous EEG decoding methods and the other NIRS-EEG decoding method, but it could also reveal reasonable activation patterns on the cortex by avoiding the contamination of putative task-irrelevant activations, which allowed us to check the validity of obtained neuro-decoders. We used simultaneous measurement of NIRS-EEG, which has the advantage of non-invasiveness, portability, and measurability in real environments. Developments in this field potentially open up possibilities to investigate brain activities in naturalistic conditions, and hence are expected to impact on the community of real-world neuroscience.

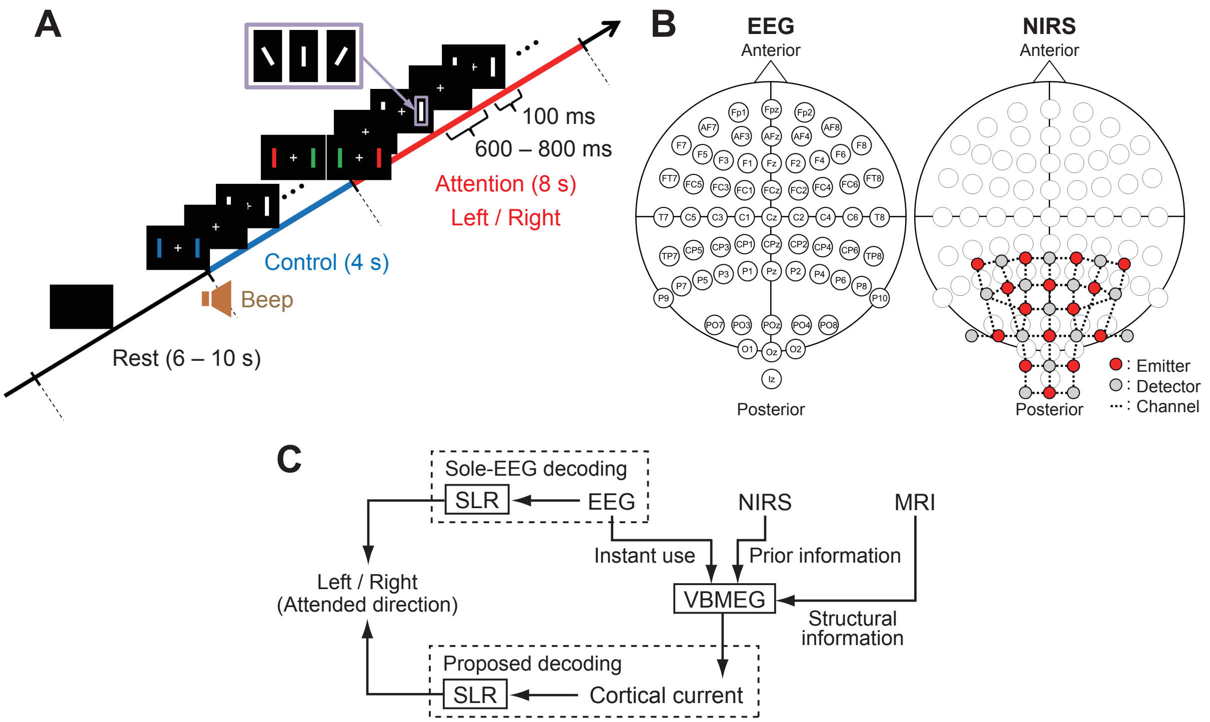


Fig. 3.1 Experimental settings and decoding concepts. (A) Timeline of the experimental block. One block consisted of 2 epochs: Control and Attention. After the onset of a block, two white flashing bars were presented repeatedly in a rapid stream at the left and right until the end of the block. During the Attention epoch, subjects were instructed to fixate their eyes to the fixation point and to covertly and continuously orient their attention to a single bar that was cued by a target color at the initial presentation of the bars. The target direction (Left or Right) was pseudo-randomly selected for each Attention epoch. The Control epoch was inserted to cancel out the difference in visual stimuli before and after the onset of the Attention epoch, thus focusing only on the modulation caused by selective spatial attention. During the Control epoch, subjects were instructed to fixate their eyes to the fixation point and distribute their attention continuously and evenly over two bars. We conducted 8 sessions for each subject. Each session consisted of 24 blocks (12 lefts and 12 rights), with short Rest epochs in between. (B) Channel configurations of EEG and NIRS. Left: Placement of EEG sensors: an International 10-20 system modified for NIRS-EEG simultaneous measurement. Right: NIRS probe positions. Emitters and detectors are indicated by red and gray circles, respectively. NIRS channels are shown by dotted lines. (C) Comparison of the concepts for the proposed method and the conventional EEG only decoding method. In the proposed method, subjects' attended direction was decoded from cortical currents estimated from EEG sensor signals with structural magnetic resonance imaging (MRI) data and an NIRS prior, where NIRS data was used as hierarchical prior information on the source current variance. The EEG only decoding scheme directly decodes attended direction from EEG sensor signals.

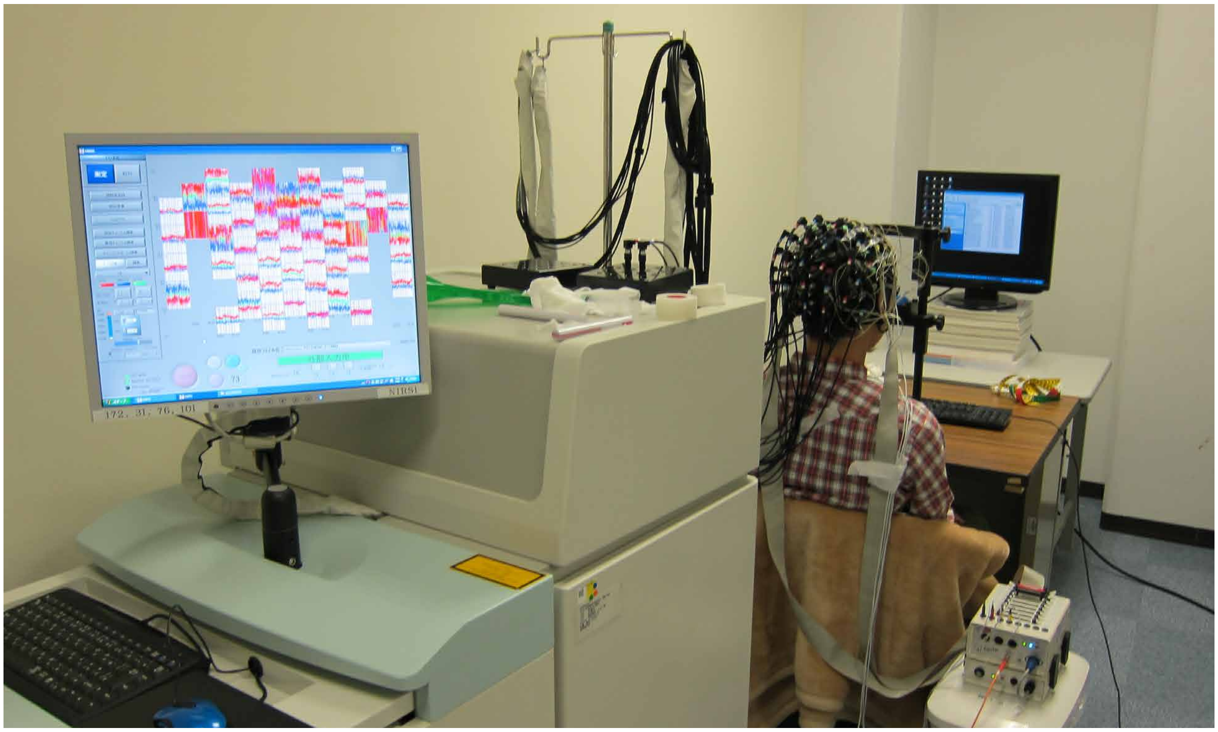


Fig. 3.2 The experimental environment used in this study for recording NIRS and EEG sensor signals simultaneously.

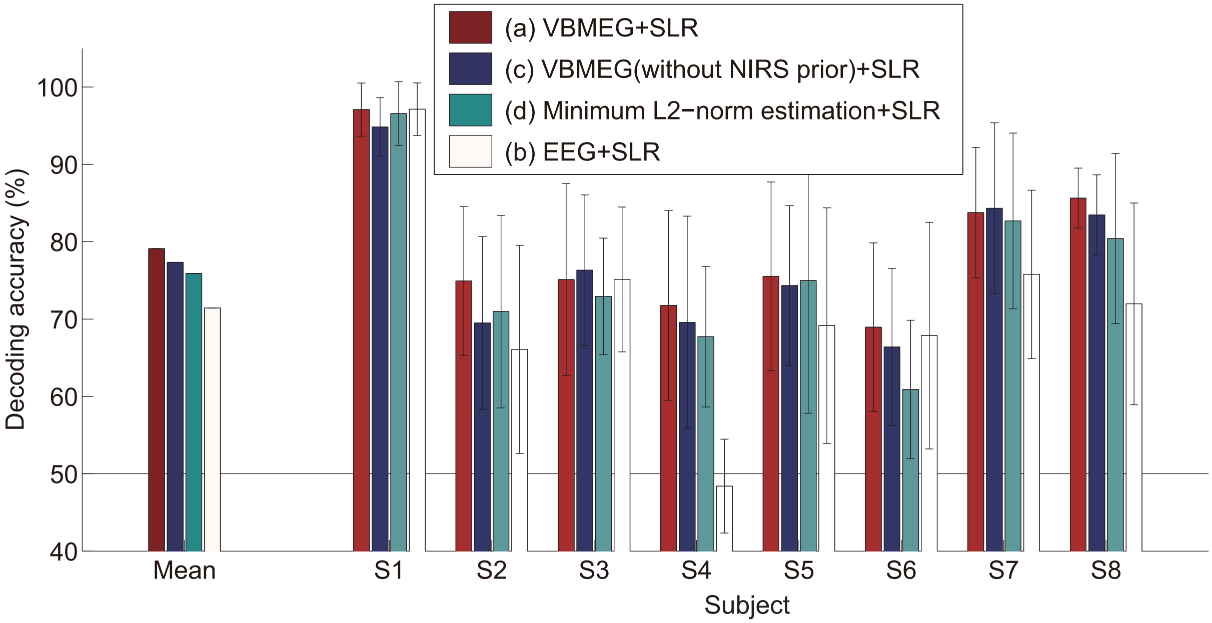


Fig. 3.3 The nested cross-validation (NCV) decoding accuracies based on (a) cortical currents estimated by VBMEG (red), (b) EEG sensor signals (white), (c) cortical currents estimated by variational bayesian multimodal encephalography (VBMEG) without NIRS prior (blue), and (d) cortical currents estimated by the minimum L2-norm method (cyan). The leftmost quadruplet of bars denote the NCV accuracies (79.1%, 77.3%, 75.9%, and 71.4%). The other bars denote NCV decoding accuracies for each subject (S1–S8). Standard deviations are denoted as the error bars. Chance level performance is indicated by the horizontal line (50%; not corrected for trial rejection).

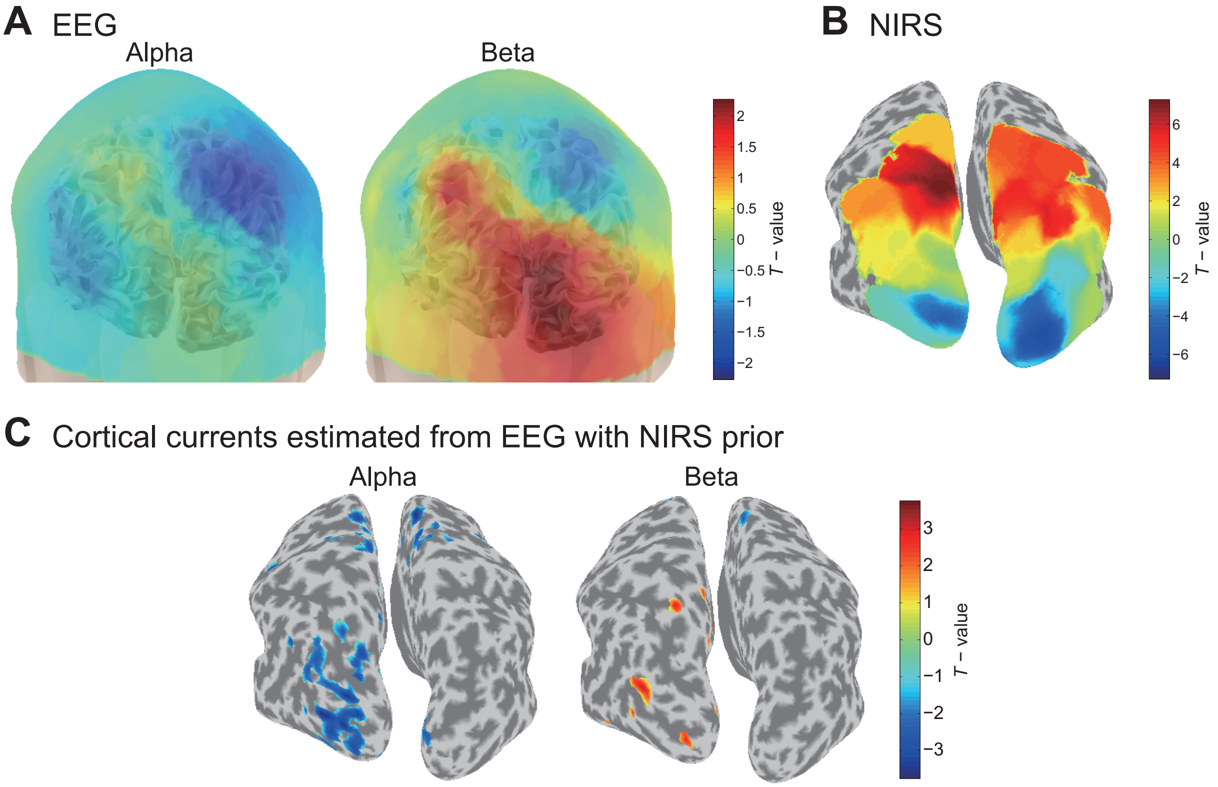


Fig. 3.4 EEG sensor signals, NIRS activities, and estimated cortical currents of S2. (A) The spatial pattern of the t-values (Attend right − Attend left) computed from EEG signals, is separately shown for two different frequency bands (alpha and beta bands), viewed from the posterior, and rotated slightly to enhance visibility. The pattern is superimposed on the subject's polygon cerebral cortex and scalp models. (B) The cortical pattern of NIRS activities. The activities are shown as t-values (Attend − Control), and rendered on an inflated cortical surface. (C) The cortical pattern of the estimated currents by VBMEG. Displayed are t-value thresholds at $p < 0.05$ (Attend right − Attend left; uncorrected) for two different frequency bands (alpha and beta bands). Sulci (concave) and gyri (convex) are indicated by dark and light-gray shading, respectively.

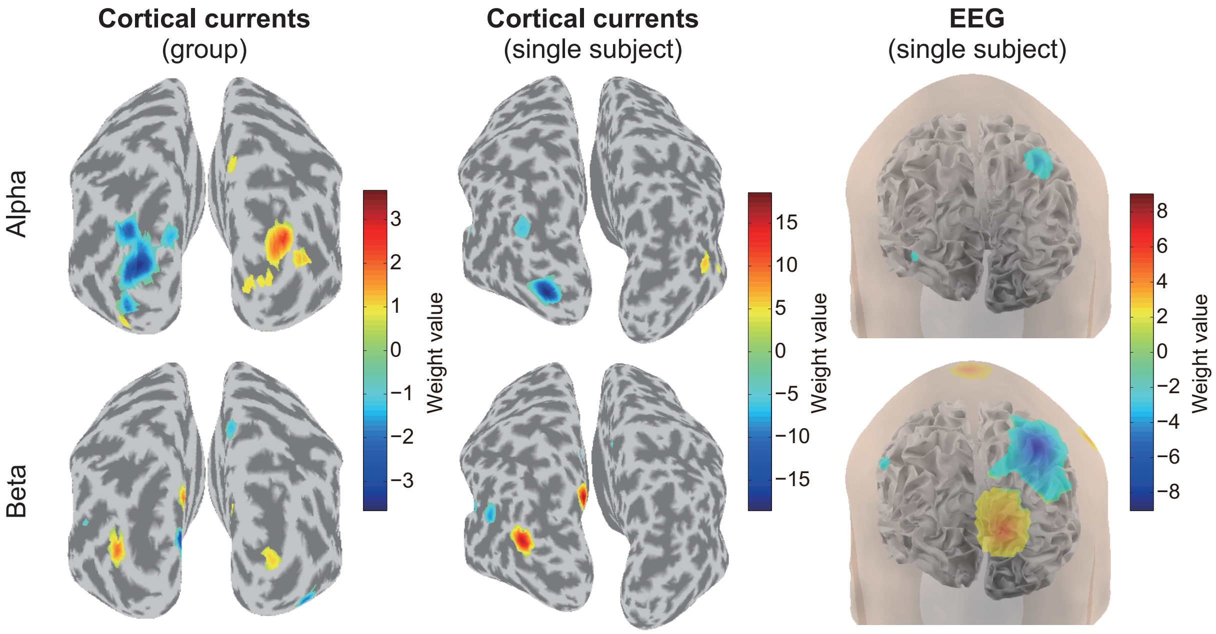


Fig. 3.5 Weights of the classifiers. Right: Weight values for EEG sensors in the classifier decoding attended direction from sensor signals in the same subject as in Fig. 3.4. These values were averaged across training sets and rendered on the scalp surface; these are shown for two different frequency bands (alpha and beta bands). Signed values of weights are represented by pseudo-color, and weight values whose absolute value is smaller than 25% of the maximum absolute value are not displayed for each panel. Center: Weights for the cortical currents in the classifier decoding attended direction from cortical currents in the same subject as in Fig. 3.4. Averaging, coloring, and thresholds are the same as the left panel, and rendered on the inflated cortical surface. Left: Weights for the cortical currents, averaged across all of the subjects. Weight values for the cortical currents for each subject were projected to a template brain and averaged. Coloring and thresholds are the same as the panels on the left, and rendered on the inflated cortical surface of the Montreal Neurological Institute template brain.

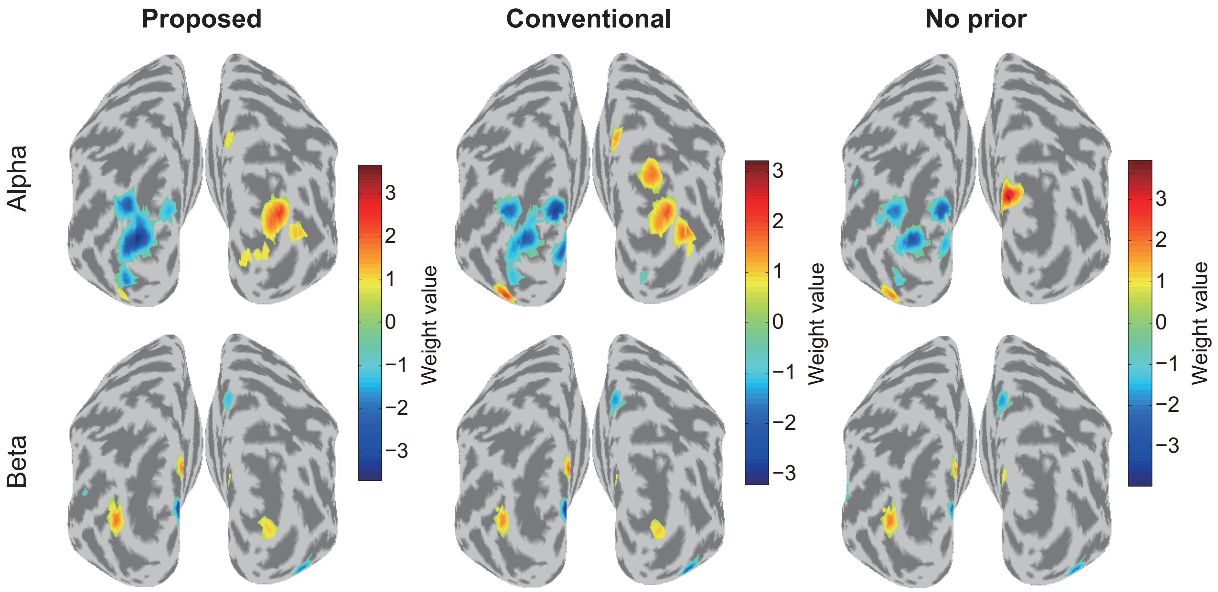


Fig. 3.6 Classifier weights. Left: Weights on cortical currents estimated with the proposed mean prior variance model, Eq. (3.2). Weights were first averaged across training sets for each subject, then projected to a template brain, and finally averaged over all of the subjects on the template. This operation was performed separately for the alpha and beta bands. Weights whose absolute value were smaller than 25% of the maximum absolute value were not rendered. Center: Cortical patterns of weights on cortical currents with the conventional mean prior variance model, Eq. (3.1). Right: Cortical patterns of weights on cortical currents with the no prior variance model.

Chapter 4

Learning a common dictionary

4.1 Introduction

One major issue in neural decoding for BMI (Dornhege et al., 2007; Graimann et al., 2011; Tan and Anton, 2010; Wolpaw and Wolpaw, 2012), as well as neuroscience studies (Haxby et al., 2001; Haynes and Rees, 2006; Horikawa et al., 2013; Kamitani and Tong, 2005; Shibata et al., 2011) is how to deal with undesired variability among different subjects or different recording sessions from a single subject. Brain signals measured over a series of experiments have inherent variability because of different physical and mental conditions among multiple subjects and sessions. Such variability complicates the analysis of data from multiple subjects and sessions in a consistent way, and degrades the performance of subject-transfer decoding in a BMI.

Several researchers (Devlaminck et al., 2011; Fazli et al., 2009; Kang and Choi, 2014; Lotte and Guan, 2010; Samek et al., 2014) tackled the subject-to-subject variability of EEG to achieve *subject-transfer decoding*, the goal of which is to classify the mental state of a previously unseen *target subject* based on the data or pre-trained decoders of other *training subjects*, so that BMI is instantly usable. There are two main approaches. First, the subject-invariant approach (Fazli et al., 2009; Samek et al., 2014) builds a universal decoder that is constructed only with data from training subjects, ignoring the specificity of target subjects. It thus reconciles to suboptimality if the target subject is dissimilar to any of the training subjects. Second, the task-calibration approach (Devlaminck et al., 2011; Kang and Choi, 2014; Lotte and Guan, 2010) conducts experiments with the target subject to obtain a task-based calibration dataset, which is used for tuning the decoder. This approach can accommodate subject-specific variation, but acquiring the task-based calibration data is often too costly in practice, especially in daily-life applications of BMI. A third approach is thus needed, which explicitly considers the variability of target subjects but is applicable based on small efforts by them.

To resolve these difficulties, we develop a novel framework for analyzing multisubject EEG data using the unsupervised signal processing technique of *dictionary learning* after compensating for variations between subjects and sessions, and design low-cost calibration for subject-transfer decoding using resting-state data. The proposed framework decomposes multichannel EEG data into a subject- (and session-) invariant dictionary of spatial pattern bases, subject- (and session-) specific linear transforms to adjust the dictionary to each subject (and session), and sparse codes. The subject-session-specific transforms are newly introduced to the dictionary learning framework to modulate the dictionary to allow compensating for variability across subjects and sessions. Dictionary learning is useful in its own right, e.g., for clearer visualizations of spatial patterns to disambiguate their neurophysiological interpretations (Barthélemy et al., 2013; Chevallier et al., 2014), and recently used for neural decoding (Hammer et al., 2011; Zhou et al., 2012). Kang and Choi (2014) has recently proposed the idea of using a latent subspace shared across subjects and tuning it with subject-specific transforms based on a Bayesian probabilistic model. We then develop a novel subject/session-transfer scheme, which uses resting-state brain activities as calibration data. The resting-state data are easy to collect, and the proposed scheme does not require expensive task-based calibration, which would be beneficial for subjects to easily use BMI. The underlying idea is that resting-state brain activity reflects the subject-specific nature of brain activity that is consistent over subsequent task sessions. Recent studies have revealed that brain activity during resting states exhibits quite organized and stable patterns of functional connectivity, such as the default mode and the dorsal attention networks (Brookes et al., 2011; de Pasquale et al., 2012; Fox et al., 2005), and is likely intrinsic to individual brains (Baldassarre et al., 2012; Massar et al., 2014; Mueller et al., 2013; Wu et al., 2014). We thus make use of resting-state data to extract subject-specific characteristics that are intrinsic and specific to individual brains, and supposed to vary more between different brains than in the same brain between different sessions. To our knowledge, the present study is the first to apply resting-state EEG data to BMI applications.

The contribution of this study is threefold.

1) We develop a new dictionary learning technique for extracting common spatial bases while compensating for variability across subjects/sessions (Fig. 4.1A).
2) We propose the use of resting-state data for calibration in subject-transfer decoding, which is made possible with the proposed dictionary learning technique (Fig. 4.1B).
3) By using real EEG recordings from more than forty subjects performing a selective visual-spatial attention task (see previous chapter), we validate the proposed dictionary learning technique, estimating interpretable spatial patterns that are consistent with existing neuroscience knowledge, and also show that the proposed subject-transfer decoding framework performs better than existing decoding methods.

4.2 Method and material

Figure 4.1 shows an overview of the proposed method described in this subsection.

4.2.1 Three core assumptions

The proposed method is based on the following three core assumptions:

A1) At each time point, brain activities as a spatial pattern can be expressed as a combination of a small number of spatial bases common across subjects and sessions.
A2) Actual signals measuring the brain activities are deformed by subject-session-specific spatial transforms.
A3) For the same subject, spatial transforms are consistent over task sessions and preceding resting-state sessions.

Figure 4.2 depicts the outline of the proposed method built based on these assumptions.

4.2.2 Requirements for data

To accomplish our goal, we require that a dataset satisfies the following three properties, which correspond to the three assumptions above, respectively:

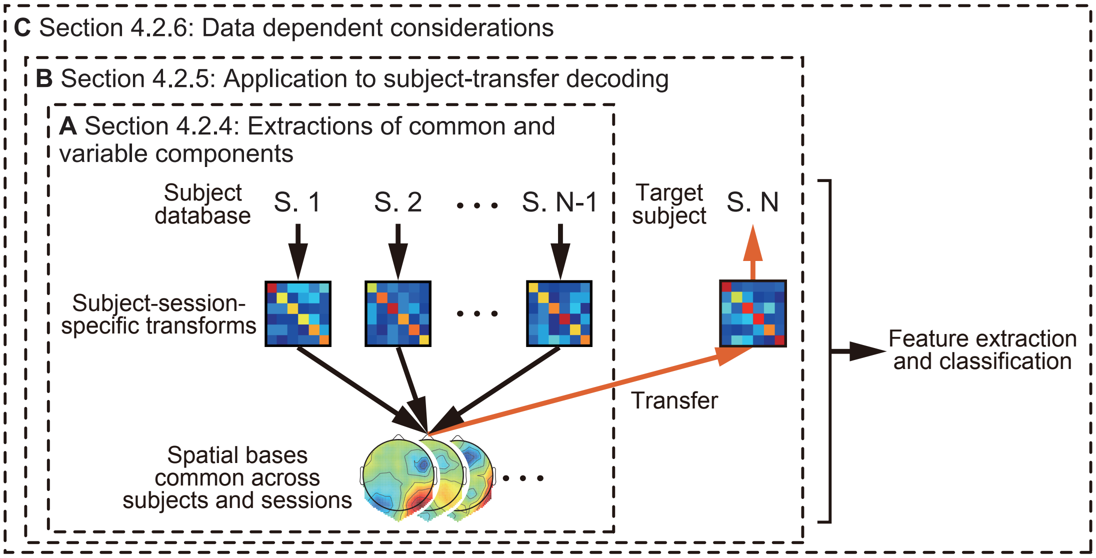


Fig. 4.1 The overview of the proposed method described in Section 4.2. The proposed method can be divided mainly into two parts; i.e., the novel dictionary learning method for extracting spatial bases common across subjects and sessions after compensating variability by simultaneously estimating subject-session-specific spatial transforms (A: Subsection 4.2.4), and its application to subject-transfer decoding with resting calibration (B: Subsection 4.2.5). Note that since the proposed dictionary learning method (A) is unsupervised, a subject-session-specific transform can be learned from resting-state data of the target subject. Subsection 4.2.6 (C) describes data dependent considerations including feature extractions and classifications.

R1) All data samples share common underlying activities; that is, all the subjects perform the same mental task such as selective spatial attention or sensorimotor rhythm modulation.

R2) Yet, the data generation process may vary over subjects and sessions due to, e.g., variations in the brain structure, differences in channel positions, changes of the conductance of skin and gel, and slight differences of brain regions activated by the same task between subjects and even sessions of the same subject.

R3) Recordings of resting-state activities of the target subjects just before performing BMI task sessions are available.

4.2.3 Basic dictionary learning

Dictionary learning is a data analysis method that estimates overcomplete bases for sparsely representing measurable signals. It has its origin in neuroscience under the name of sparse coding (Olshausen and Field, 1997), which is still a current research topic (Hunt et al., 2013), and has been applied to signal processing (Elad, 2010; Mallat, 2008; Patel and Chellappa, 2013) for denoising, compression, and so on. Sparse representation has also been shown to improve classification performance in pattern recognition (Gao et al., 2010; Mairal et al., 2008; Zhang and Li, 2010). Instead of using predefined bases like discrete cosine transform (DCT) basis or wavelets, dictionary learning adaptively constructs a set of bases, or a *dictionary*, from the given data with sparseness constraints, so that the dictionary is best suited for representing the data at hand. Principal component analysis (PCA) also estimates an orthogonal bases from the given data; however, the sparse method is more flexible as its overcomplete basis can cover some dynamic characteristics that may be possessed by many real-world signals.

We satisfy the first assumption A1) by dictionary learning. According to the basic formulation of dictionary learning, a vector of measured signals[*1] $\mathbf{x}_t \in \mathbb{R}^M$ at time t is represented by

$$\mathbf{x}_t \approx \mathbf{D}\boldsymbol{\alpha}_t, \tag{4.1}$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \mathbb{R}^{M \times K}$ is a dictionary matrix whose column vectors $\mathbf{d}_k$ are called atoms and $\boldsymbol{\alpha}_t \in \mathbb{R}^K$ is called a sparse code. This equation can be seen as a conversion from a signal $\mathbf{x}_t$ to a sparse code $\boldsymbol{\alpha}_t$. If $\mathbf{D}$ were fixed at a DCT basis, then $\boldsymbol{\alpha}$ would be a frequency domain variable. Dictionary learning estimates $\mathbf{D}$ adaptively based on the given data. In general, we take $K > M$ (more bases than the signal dimensionality), that is, we use an “overcomplete” dictionary $\mathbf{D}$. The dictionary is estimated from the given data by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{A}} \frac{1}{2} \sum_t \|\mathbf{x}_t - \mathbf{D}\boldsymbol{\alpha}_t\|_2^2 \\ &\text{s.t. } \forall t\colon \|\boldsymbol{\alpha}_t\|_0 \le L,\ \forall k \in \{1, \ldots, K\}\colon \|\mathbf{d}_k\|_2 = 1, \end{aligned} \tag{4.2}$$

where $\mathbf{A} \equiv [\boldsymbol{\alpha}_t]$ is a set of sparse codes, $\|\cdot\|_0$ is the ℓ_0-norm, and $\|\cdot\|_2$ is the ℓ_2-norm. Each sparse code $\boldsymbol{\alpha}_t \in \mathbb{R}^K$ is constrained to have L or less non-zero elements. The unit ℓ_2 norm constraint for $\mathbf{d}_k$ is to eliminate the scale indeterminacy between $\mathbf{D}$ and $\boldsymbol{\alpha}$ (actually, the product of scaled variables $(s\mathbf{D})(s^{-1}\boldsymbol{\alpha})$ would give the same $\mathbf{x}$ for any $s \neq 0$). Solving this unsupervised reconstruction problem, we obtain the dictionary (and the sparse codes) that best represents the

---

[*1] The vector $\mathbf{x}$ may be direct measurements, or signals preprocessed depending on the characteristics of the data. We will revisit the preprocessing in Section 4.2.6.

given signals. In the context of our EEG neuroimaging, M is the number of channels spread over the scalp and each element of the signal vector $\mathbf{x}$ is the measurement at the corresponding channel. The learned dictionary $\mathbf{D}$ is then interpreted as the bases of spatial activity patterns (i.e., spatial bases). With the sparsity assumption of $\boldsymbol{\alpha}_t$, EEG signals $\mathbf{x}$ are represented by a linear combination of sparsely selected spatial bases, and this is what our first assumption A1) has stated.

The variability in the signal measurement is ignored in the basic formulation of dictionary learning, Eq. (4.1), which deals with the variety in signal patterns by means of an overcomplete basis. That is, variations in the measured signal $\mathbf{x}_t$ can only be explained in the variety in the bases (atoms) in the dictionary. If the dataset contains measurements with different channel alignments on the scalp, the dictionary needs to have an individual set of bases for each of the different physical settings, even for representing the same underlying brain activities. As a result, the dictionary can be too redundant and lack generalization capability. A similar situation may occur when brain activities involve non-stationarity. This in turn makes sparse representation unstable and will result in decreased classification performance, which poses a challenge in developing neuro-decoders that generalize over different subjects and/or different sessions.

4.2.4 Proposed dictionary learning for common and variable components

To resolve the problem in the basic formulation of dictionary learning, we adopt the second assumption A2) and propose a novel formulation, which includes subject-session-specific transformation of the dictionary (Fig. 4.2A). To accommodate the variability over different settings, we introduce a spatial transform that modulates the common dictionary. The transform depends on subjects and sessions so as to be responsible for compensating for the differences in subjects and sessions. Note that if we do not consider subject-session-specific spatial transforms (or employ an identity mapping for the transform), the proposed formulation reduces to the original formulation in Eq. (4.1).

Based on assumptions A1) and A2), we re-formulate Eqs. (4.1) and (4.2). In our pro-

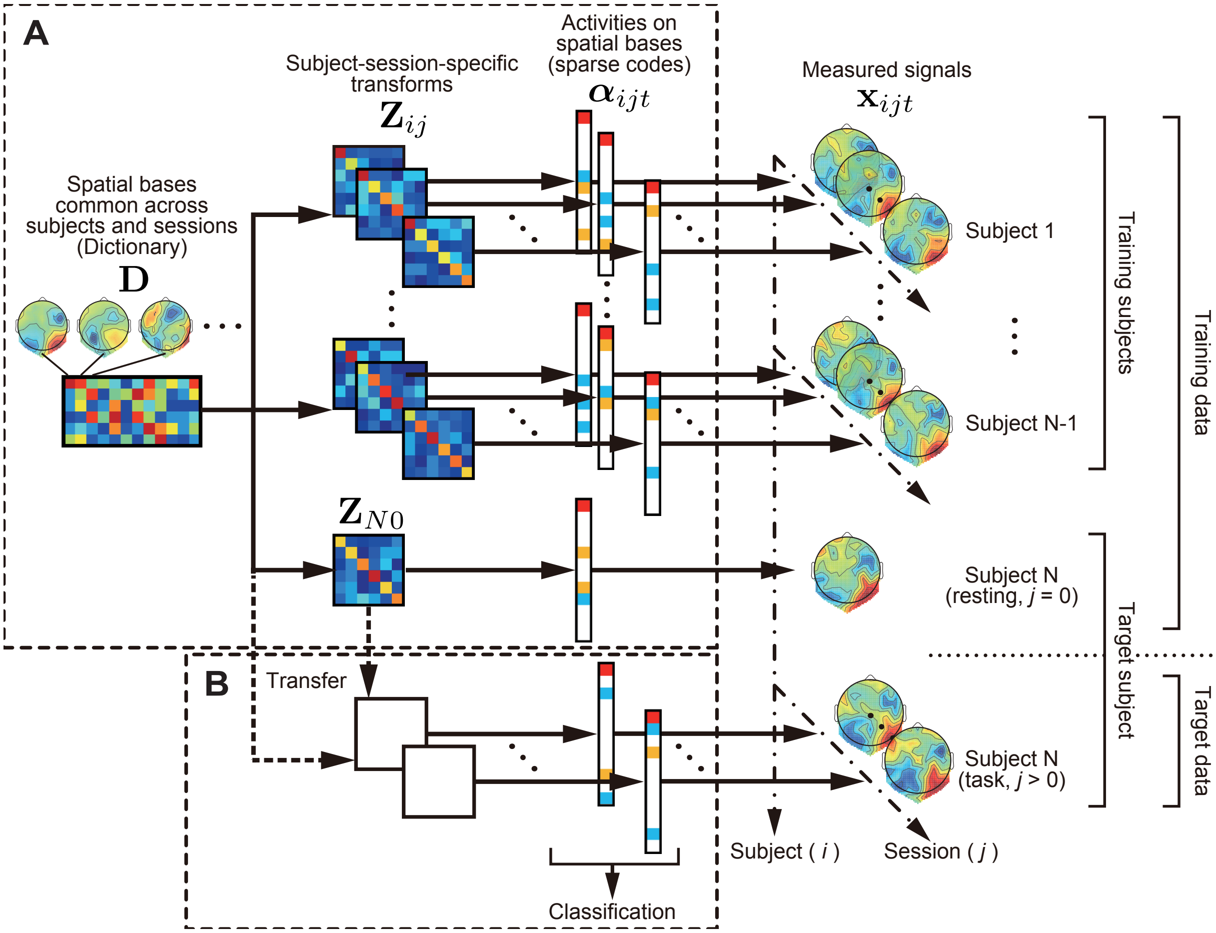


Fig. 4.2 The proposed method for multi-subject-session data analysis and subject-transfer decoding. (A) Measured signals are decomposed into three factors: dictionary (spatial bases) common across subjects and sessions $\mathbf{D}$, spatial transforms $\mathbf{Z}_{ij}$ that modulate the common dictionary to be specific to subjects (indexed by i) and sessions (indexed by j), and sparse codes $\boldsymbol{\alpha}_{ijt}$ representing the use profile of the spatial bases. The dictionary and transforms are learned from the training data consisting of all sessions of training subjects and a resting session (indexed by $j = 0$) of a target subject (indexed by $i = N$), in an unsupervised manner so that the signals in the training data can be well represented by a sparse combination of the spatial bases in the dictionary. (B) Subject-transfer decoding based on calibration using resting-state activities. For the target subject ($i = N$), the subject-session-specific transform for the resting-state activities, $\mathbf{Z}_{N0}$, is re-used as those for the task sessions, $\mathbf{Z}_{Nj}$ ($j > 0$). The two matrices $\mathbf{D}$ and $\mathbf{Z}_{N0}$ are obtained in the dictionary learning stage of (A) based on the training data, and are thus available before getting access to the EEG measurements of the target subject during the decoding task sessions. Using the EEG signals of the target subject during the decoding task, sparse codes $\boldsymbol{\alpha}$ are estimated and used for later classification.

posed method, signal measurements are given by

$$\mathbf{x}_{ijt} \approx \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}_{ijt}, \tag{4.3}$$

where i, j, t are indices for subjects, sessions, and time points, respectively, $\mathbf{x}_{ijt} \in \mathbb{R}^M$ is a vector of measured signals at (i,j,t), $\mathbf{Z}_{ij} \in \mathbb{R}^{M\times M}$ is a matrix of a spatial transform specific to the subject-session pair (i,j), $\mathbf{D} \in \mathbb{R}^{M\times K}$ is a dictionary common across subjects and sessions, and $\boldsymbol{\alpha}_{ijt} \in \mathbb{R}^{K}$ is a sparse code for $\mathbf{x}_{ijt}$. Whereas the dictionary $\mathbf{D}$ is unique and shared over subjects and sessions, different $\mathbf{Z}_{ij}$ are prepared for each subject-session pair. Here, $\mathbf{Z}_{ij}$ does not depend on the time point t; that is, the spatial transform may be different for each subject-session, but stays the same during a session. The transformed dictionary $\mathbf{Z}_{ij}\mathbf{D}$ serves as the set of bases for representing $\mathbf{x}_{ijt}$. In Eq. (4.3), $\mathbf{Z}_{ij}$ and $\mathbf{D}$ lack identifiability in their rotation and permutation; e.g., if two rows of $\mathbf{D}$ are swapped and the corresponding columns of $\mathbf{Z}_{ij}$ are also swapped, their product still produces the same matrix as the one without the swap. Our proposed optimization problem for estimating the transforms and the dictionary is as follows:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{D},\mathbf{A}} \frac{1}{2} \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}_i} \sum_{t\in\mathcal{S}_{ij}} \left\{ \|\mathbf{x}_{ijt} - \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}_{ijt}\|_2^2 + \lambda \|\mathbf{Z}_{ij} - \mathbf{I}\|_{\mathrm{F}}^2 \right\} \\ &\text{s.t. } \forall i,j,t\colon \|\boldsymbol{\alpha}_{ijt}\|_0 \le L,\ \forall k \in \{1,\ldots,K\}\colon \|\mathbf{d}_k\|_2 = 1, \end{aligned} \tag{4.4}$$

where $i \in \mathcal{S}$ indexes the subjects, $j \in \mathcal{S}_i$ indexes the sessions performed by subject i, $t \in \mathcal{S}_{ij}$ indexes the time points in session j by subject i, $\mathbf{Z} \equiv [\mathbf{Z}_{ij}]$ is a set of the spatial transforms, $\mathbf{A} \equiv [\boldsymbol{\alpha}_{ijt}]$ is a set of the sparse codes, $\lambda \ge 0$ is a regularization constant, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. We also let $\mathbf{X} \equiv [\mathbf{x}_{ijt}]$ be a matrix comprising the given signals. Ideally, the variation across subjects and sessions is absorbed into the subject-session-specific spatial transforms $\mathbf{Z}_{ij}$, allowing us to estimate dictionary $\mathbf{D}$ and sparse code $\boldsymbol{\alpha}_{ijt}$ by disregarding the subject-session dependent variations.

The regularization term newly introduced in Eq. (4.4) controls to what degree each $\mathbf{Z}_{ij}$ may be deviated from the identity transform, which preserves the consistent correspondence between row vectors in the dictionary and sensor positions over subjects, and also reduces the inherent indeterminacy of $\mathbf{Z}_{ij}$ and $\mathbf{D}$. In this regularization, we prefer a spatial transform ($\mathbf{Z}_{ij}$) which is not distant from the identity transform, assuming that spatial variability among subjects and sessions is not very large, so that the role of the spatial transform is to slightly distort the subject-session-independent dictionary. This assumption is realistic since the shapes of heads or brains share some degree of similarity among subjects, and we usually try to keep the channel positions consistent over sessions according to the 10-20 system, even if those cannot be completely the same. Importantly, the regularization allows us to interpret the dictionary atoms

```
Input: $\mathbf{X}$ (given signals), $K$ (dictionary size), $L$ (sparseness constant), $\lambda$ (regularization constant), $R$ (number of iterations), and $\rho$ and $r_0$ (learning rate constants)
1: Initialize $\{\mathbf{Z}_{ij}\}$ and $\mathbf{D}$ (see text)
2: for $r = 1$ to $R$ do
3:     Draw a pair of indices (subject, session) $= (i, j)$
4:     Draw a time index $t$ for $(i, j)$ and $\mathbf{x}_{ijt}$ from $\mathbf{X}$.
5:     Sparse coding: Compute $\boldsymbol{\alpha}^\star$ using OMP:
           $\boldsymbol{\alpha}^\star \leftarrow \operatorname*{argmin}_{\boldsymbol{\alpha}} \|\mathbf{x}_{ijt} - \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}\|_2^2 \quad \text{s.t. } \|\boldsymbol{\alpha}\|_0 \leq L$
6:     Choose the learning rate $\rho_r \leftarrow \rho \min(1, r_0/r)$.
7:     Update the matrices by a gradient descent:
       $\mathbf{D} \leftarrow \mathbf{D} - \rho_r(-\mathbf{Z}_{ij}^{\mathrm{T}}(\mathbf{x}_{ijt} - \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}^\star)\boldsymbol{\alpha}^{\star\mathrm{T}})$,
       $\mathbf{Z}_{ij} \leftarrow \mathbf{Z}_{ij} - \rho_r(-(\mathbf{x}_{ijt} - \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}^\star)(\mathbf{D}\boldsymbol{\alpha}^\star)^{\mathrm{T}} + \lambda(\mathbf{Z}_{ij} - \mathbf{I}))$.
8:     Normalize the dictionary atoms to make their $\ell_2$ norms unity.
9: end for
10: Sparse coding: Compute $\mathbf{A}$ using OMP for all $(i, j, t)$ with Step 5.
11: return $\mathbf{Z}$, $\mathbf{D}$, and $\mathbf{A}$
```

Fig. 4.3 A stochastic gradient descent algorithm for the optimization problem Eq. (4.4).

as activity patterns on the channel positions for all the subjects. Such an interpretation cannot be made, e.g., if $\mathbf{Z}_{ij}$ permutes the rows of dictionary matrix $\mathbf{D}$; our regularization prevents $\mathbf{Z}_{ij}$ from being close to any non-identity permutation matrix. Another important consequence of this regularization is that it reduces the indeterminacy of $\mathbf{Z}_{ij}$ and $\mathbf{D}$: Given a solution $([\mathbf{Z}_{ij}^*], \mathbf{D}^*)$ of the problem (4.4), $([\mathbf{Z}_{ij}^*\mathbf{R}^{\mathrm{T}}], \mathbf{R}\mathbf{D}^*)$ with any orthogonal matrix $\mathbf{R}$ also gives another solution if $\lambda = 0$, but this does not hold if $\lambda > 0$. The regularization thus removes the indeterminacy of rotation in our model. Notice also that if every $\mathbf{Z}_{ij}$ is set to the identity matrix (i.e., $\lambda \to \infty$), the problem (4.4) reduces to the basic one (4.2). Our method thus includes the basic dictionary learning as a special case by not allowing any variability across subjects and sessions.

Figure 4.3 shows a stochastic gradient descent algorithm (Mairal et al., 2012) for solving the constrained optimization problem Eq. (4.4). Since the optimization problem is not convex, the solution quality depends substantially on the initial setting of the algorithm. The dictionary $\mathbf{D}$ was initialized as a solution of the basic dictionary learning problem, Eq. (4.2), by running the SPAMS toolbox (Mairal et al., 2010). Although the basic dictionary learning problem is also non-convex, the stochastic gradient descent algorithm for it has empirically been found to be stable regardless of an initialization (Mairal et al., 2012, 2010); in our initialization, we used a dictionary whose atoms were randomly drawn samples from a training dataset. The initial values for $\mathbf{Z}_{ij}$ were obtained by ridge regression, fixing $\mathbf{D}$ and $\mathbf{A}$ at the solution of Eq. (4.2), and then by scalar scaling to minimize $\|\mathbf{Z}_{ij} - \mathbf{I}\|_{\mathrm{F}}$. After the initialization, the stochastic gradient descent algorithm for our optimization Eq. (4.4) (Fig. 4.3) was executed. In the "for" loop of the algorithm, we draw one combination of a subject and a session (i.e., i and j) in lexicographic order, and then draw a signal vector $\mathbf{x}_{ijt}$ from the selected (i, j) from randomly permuted within-session time indices. For the signal vector $\mathbf{x}_{ijt}$, the sparse code $\boldsymbol{\alpha}^{\star}$ was computed by the orthogonal matching pursuit (OMP) algorithm (Mallat and Zhang, 1993) so that $\mathbf{x}_{ijt} \approx \mathbf{Z}_{ij}\mathbf{D}\boldsymbol{\alpha}^{\star}$. After that, we updated $\mathbf{D}$ and $\mathbf{Z}_{ij}$ in the gradient descent step. The learning rate ρ_r at the rth iteration in the loop was set at a constant ρ during the first 10% iterations (i.e., $r = 1, \ldots, R/10$), and afterward it decreased in reciprocal proportion to the number of iterations. The scheduling was controlled in Step 6, where we set $r_0 = R/10$. In actual implementation, to achieve acceleration, we adopted a mini-batch strategy (Mairal et al., 2012), which was to draw many (η) time indices and signal vectors, rather than single ones, in Step 4 of the algorithm. In that case, η sparse codes were calculated from the drawn η samples (in Step 5), and gradients for $\mathbf{D}$ and $\mathbf{Z}_{ij}$ were calculated from the η sparse codes (in Step 7).

4.2.5 Application to subject-transfer decoding

The objective of subject-transfer decoding is to predict the mental state of a target subject using a classifier that has been trained by a dataset mainly from subjects other than the target subject, whereas the standard "self-decoding" scheme attempts to predict the mental state of a target subject using a classifier that has been trained by a dataset collected from the same subject performing the task sessions. When constructing a BMI decoder, there are two stages. Since brain activities are usually of high dimensionality, we first extract an essential feature vector from brain signals so as to well represent the mental state of the subject. Such a feature vector is then provided to a classifier that predicts the mental state based on the feature vector. When training a decoding classifier, we know the true mental state as experimental data, so the classifier can be trained to well associate the input feature vector with the output label both registered in the training dataset. When testing, the trained classifier is required to predict the label when provided each feature vector in the test dataset, and then its performance is evaluated based on how accurately the labels in the test dataset are predicted.

The use of the representations of brain activities on the common spatial bases for training/testing the classifier allows us to compare different brain activities on a common domain, hence leading to successful subject-transfer decoding. As the feature vector, we use sparse codes $\boldsymbol{\alpha}_{ijt}$, which are interpreted as the brain activities mapped onto the domain of spatial bases spanned by the dictionary atoms common across subjects and sessions.

Although the estimation of the sparse codes of the training subjects is straightforward from the algorithm shown in Fig. 4.3, that of the target subject needs special consideration because we do not have the data from BMI task sessions of the target subject in the training stage; this is a natural consequence of the problem setting of subject-invariant decoding. This means that we cannot prepare the transforms specific to the target subject in the training stage, while those transforms are necessary to estimate sparse codes of the task sessions of the target subject. Since we compute feature vectors provided to the classifier from sparse codes, the lack of these transforms makes the BMI/decoding scenario collapse.

To cope with this, we use resting session data from the target subject, which are assumed to be available in the training stage, and use the transform $\mathbf{Z}_{N0}$ for the resting session ($j = 0$) of the target subject ($i = N$) as a substitute for $\mathbf{Z}_{Nj}$ $(j > 0)$. The transform $\mathbf{Z}_{N0}$ is estimated simultaneously with those of all sessions of the other training subjects. This operation is simply performed by applying our dictionary learning algorithm (Fig. 4.3) to the dataset of the training subjects added by the resting session data of the target subject (Fig. 4.2A). In each trial in the target subject's task session (indexed by j (> 0)), we can immediately compute the sparse code $\boldsymbol{\alpha}_{Njt}$ by OMP, Step 5 in Fig. 4.3, by using the transform $\mathbf{Z}_{N0}$ instead of unknown $\mathbf{Z}_{Nj}$ $(j > 0)$, and the subject/session independent dictionary $\mathbf{D}$. See Fig. 4.2B for a schematic diagram. The basic idea of our subject-transfer decoding is to transfer the subject/session independent dictionary $\mathbf{D}$ into the subject-session dependent dictionary $\mathbf{Z}_{N0}\mathbf{D}$, by using the spatial transform calibrated with the resting-state activities, $\mathbf{Z}_{N0}$. This transfer of the resting session's transform to the task session's transforms is based on the third assumption A3); the transform estimated from the resting session data is assumed to be comparable to those estimated from the task session data. The brain activities of the target subject in the task sessions are decoded by applying the estimated sparse codes to the classifier trained based on the task sessions performed by the training subjects.

Since how to convert sparse codes to features supplied for the classifier depends on the characteristics of the data and the task, we devote the following subsection to describing an evaluation dataset and feature extraction procedures.

4.2.6 Data-dependent considerations: Experimental procedure and feature extraction

We evaluated the proposed method both for extracting spatial bases common across subjects and for subject-transfer decoding using the settings described below.

Experimental procedure

Fifty-one volunteers participated in the experiments, and we used only data of 45 out of 51 subjects who performed both resting and task sessions without experimental or measurement instruments' problems. All the subjects gave written informed consent for an experimental procedure described below, which had been approved by the ATR Human Subject Review Committee.

In the task sessions, the subjects performed a selective visual spatial attention task (attend-left or attend-right); the subjects were requested to covertly attend to left or right following visual stimulus cues without moving their eyes.

The experimental procedures were almost the same as that used in the previous chapter except for the acquisition of resting-state activities. During the first (0th) session for each subject, resting-state activities were measured for five minutes; the subjects were requested to stay awake, relax, keep their body still, fixate their open eyes on a central white cross, and not concentrate on a specific thought. After the resting session, each subject performed eight (1st to 8th) task sessions, each consisting of 24 blocks. Thus, the total number of trials (Attention epochs) was 192 for each subject.

Data acquisition and preprocessing

In this chapter, we assumed that only EEG measurements were available for analyses, which is a natural situation for the practical BMI. Although EEG has advantages of low cost and low burden, it has a disadvantage of large variability across subjects and sessions. In addition, in this case the cortical current estimation is not available because we need NIRS and MRI data in addition to the EEG data for a reliable cortical current estimation (Chapter 3). That means whether we perform the cortical current estimation or not depends on a tradeoff between costs for data acquisition and a quality of signals.

EEG signals were measured using an ActiveTwo system by Biosemi (Amsterdam, the Netherlands), at a sampling rate of 256 Hz with a 64-electrode cap that was configured according to the international 10-20 system. We also measured EOG to ensure that the subjects did not shift their eye gaze towards the attended direction by checking if the attended direction could not be decoded only from EOG (see the first paragraph of Section 4.3.2).

We preprocessed the raw EEG signals to obtain the input, $\mathbf{X}$, to the dictionary learning algorithm. First, the raw EEG signals were re-referenced to a common average reference and passed through a band-pass filter (0.5–40 Hz). Next, we derived the spectral power in seven frequency bands using Morlet wavelets. We chose a spectral standard deviation of 5.83 and spaced the center frequencies evenly in the logarithmic scale from 2^2 to 2^5 with a step of 0.5, that is, $[2^2, 2^{2.5}, 2^3, 2^{3.5}, 2^4, 2^{4.5}, 2^5] \approx [4, 5.7, 8, 11.3, 16, 22.6, 32]$. We used the $F = 7$ frequency bands based on the knowledge that the selective spatial attention task modulates a wide range of frequency bands (Daitch et al., 2013; Siegel et al., 2008); this knowledge has also been used in other BMI studies to enhance decoding performance (Ang et al., 2008). After the wavelet transform, the data were downsampled to 32 Hz and normalized to have zero mean and unit ℓ_2 norm for each time point and frequency, leading to $\mathbf{X} = [\mathbf{x}_{ijt}]$ in Fig. 4.3. The dimensionality M of the input $\mathbf{x}_{ijt}$ was the same as the number of channels, 64, for each frequency band.

We excluded behaviorally poor trials from the later analysis with the same criterion as in the previous chapter, which rejected trials that were not significant with the significance level of 0.05 under the null hypothesis that the rate of correct responses was comparable to the chance level. In addition, we excluded subjects whose trial rejection rate was more than 25%. Based on the latter criterion, four subjects were removed, and henceforth we analyzed 41 subjects.

Feature extraction and classification

The performance of subject-transfer decoding was evaluated by one-subject-out CV accuracy; that is, we picked up one subject from the 41 subjects as a target subject, constructed a decoder based on a training dataset consisting of the data from the remaining 40 subjects and the resting-state data from the target subject, evaluated the decoding accuracy by decoding the task dataset from the target subject, and repeated this by replacing the target subject 41 times. We call the decoding accuracy in one repetition the *transfer performance* for the picked target subject, and the median of the 41 transfer performances is called the *CV transfer performance*.

Classification was performed using a linear support vector machine (SVM) which has been successfully used for image classifications based on the dictionary learning and sparse coding (Wang et al., 2010; Yang et al., 2009). The SVM was trained using the task data from the training subjects, and objective of the SVM was to predict the labels of the spatial attention trials performed by a target subject.

The sparse codes averaged over each trial were used as a feature vector for training and testing the SVM classifier (Fig. 4.2). The sparse codes of the training subjects were estimated using data points within 1) Attention epochs of task sessions of the training subjects, and 2) entire intervals of resting sessions of the training and target subjects, using the proposed dictionary learning algorithm (Fig. 4.3). In the estimation procedure, when we picked up a resting session in Step 3 in Fig. 4.3 (i.e., when $j = 0$), we updated only $\mathbf{Z}_{ij}$ in Step 7 in order to make $\mathbf{D}$ specific to task sessions. After obtaining the sparse codes, we averaged the sparse codes in each trial, producing a vector of length K, and normalized it to have a unit ℓ_2 norm. We concatenated the averaged sparse codes over the $F = 7$ frequency bands to make the feature vector for the trial (the dimensionality of the feature vector is FK). The SVM was trained based on these feature vectors obtained from the training subjects, which were associated with the corresponding task label (left or right) for each trial. In each Attention epoch of task sessions of the target subject, sparse codes were extracted by using the transfer scheme described in Section 4.2.5. The feature vector for the target subject was computed in the same way as the training subjects, by taking average of the sparse codes in a trial, and supplied to the SVM trained by the dataset of the training subjects. The decoding accuracy was the proportion of correctly predicted labels for task trials of the target subject.

Parameter settings

We set the parameters in dictionary learning based on the practice in dictionary learning and some initial attempts: 1) Dictionary size $K = 128$, which was twice the feature dimensionality to keep a variety of dictionary atoms while reserving computational efficiency. 2) Sparseness constant $L = 20$, which was about one third of the feature dimensionality, reducing the number of active elements to lead to sparse solutions while keeping enough information for decoding. 3) Regularization constant $\lambda = 10^{-7}$, which well balanced the stability and adaptability of $\mathbf{Z}_{ij}$. Although these parameters were chosen based on some initial attempts in this study, these values can be chosen based on a reconstruction error or a nested-CV scheme. Other parameters in the optimization procedure were determined based on the convergence profile of the objective function, Eq. (4.4), during the stochastic gradient descent algorithm: Number of iterations $R = 6 \times 10^5$, initial learning rate $\rho = 5$, and mini-batch size 512. A hyperparameter of the linear SVM was determined without using the target subject's data by nested CV; that is, by running a CV procedure for each of the 41 training datasets.

4.3 Results

We examined the proposed method in terms of the following two aspects: 1) capability to extract common bases and 2) applicability to subject-transfer decoding.

4.3.1 Spatial bases common across subjects and sessions

We examined how the variability across different subjects and sessions was compensated for by our dictionary learning method. We show here analysis for the alpha frequency band (at a center frequency of $2^{3.5} = 11.3$ Hz) because it is known to be most relevant to our selective spatial attention task (Capotosto et al., 2009; Thut et al., 2006).

The proposed formulation of dictionary learning in Eq. (4.3) extracted sparse codes $\mathbf{A}$ that were more similar across subjects and sessions, and thus were more consistent than those by the basic formulation of Eq. (4.2) (Fig. 4.4A). To see how the dictionary learning with subject-session-dependent spatial transformation works, we compared the intersession distances in the sparse codes estimated using the proposed dictionary learning with those using basic dictionary learning (Fig. 4.4A). The distance between two sessions is defined to be the cosine distance of the averaged sparse codes (i.e., for a training subject-session pair ij and a target subject-session pair $i'j'$, we took the cosine distance between $\overline{\boldsymbol{\alpha}}_{ij} = \sum_t \boldsymbol{\alpha}_{ijt}$ and $\overline{\boldsymbol{\alpha}}_{i'j'} = \sum_t \boldsymbol{\alpha}_{i'j't}$; i.e., $1 - \overline{\boldsymbol{\alpha}}_{ij} \cdot \overline{\boldsymbol{\alpha}}_{i'j'} / \|\overline{\boldsymbol{\alpha}}_{ij}\| \|\overline{\boldsymbol{\alpha}}_{i'j'}\|$). Paying particular attention to the consistency of the sparse codes, we calculated the distances of the sparse codes in task sessions between training subjects and a target subject within the CV procedure. In the CV procedure, we selected one target subject out of 41, and the remaining 40 subjects were used for dictionary learning. After learning, we measured the distance between each of the 40 pairs of a target subject and a training (reference) subject; since every subject performed eight task sessions, there were $40 \times 8 \times 8 = 2,560$ session-wise distances over all the pairs of a target subject and reference subject. The total of $41 \times 2,560 = 104,960$ distance values are shown in each of the histograms in Fig. 4.4A. Whereas basic dictionary learning produced session-wise cosine distances that were widely distributed over $[0.2, 1]$, the distance distribution of the proposed method incorporating subject-session dependent transforms $\mathbf{Z}_{ij}$ is sharply peaked at around 0.2, leading to more consistent sparse codes.

The subject-session-specific spatial transforms $\mathbf{Z}_{ij}$ obtained by our dictionary learning were similar to each other in individual subjects, regardless of whether they were from resting sessions or task sessions. Figure 4.4B is a multidimensional scaling (MDS) plot of the spatial transforms $\mathbf{Z}_{ij}$ for all the pairs of 41 subjects (indexed i) and nine sessions (indexed j). The distance for MDS was given by the Frobenius norm of the difference, ($\|\mathbf{Z}_{ij} - \mathbf{Z}_{i'j'}\|_{\mathrm{F}}$). Here we did not employ CV but used the entire data for learning. The spatial transforms formed clusters for each subject, suggesting that the intersubject variability of the spatial transforms was larger than the intrasubject (intersession) variability. Another important point is that the spatial transforms of the same subject are closely located regardless of the session type, suggesting the subject-wise consistency of the spatial transforms between resting and task sessions.

The learned dictionary $\mathbf{D}$ included spatial activity patterns relevant to the selective spatial attention task. The left panels of Fig. 4.4C show learned prototypes (atoms) that had the top four largest differences in the associated sparse codes between the two attention conditions, left and right, in the task sessions. The displayed prototypes were associated with the smallest p values based on the Wilcoxon rank sum test, for which only the sparse code time courses (right panels) from 100 ms to 500 ms after each onset of visual stimulus during the Attention epoch were used. The prototypes (spatial bases) labeled (i)–(iv) in Fig. 4.4C can be interpreted as activities in (i) the middle temporal area (MT+), (ii) the posterior intraparietal sulcus (pIPS), (iii) the FEF, and (iv) the ventral intraparietal sulcus (vIPS). The right panels of Fig. 4.4C show the temporal activity patterns of the sparse codes for the four prototypes, averaged over all the subjects for each of the two task conditions, attend left and attend right. Since the Attention epochs of 8 s were separated into "attend-left" and "attend-right" conditions, the time courses were averaged separately within each of the two attention conditions, colored orange and blue, respectively. When taking the average over the subjects, sparse code activities were time-aligned to the average times of the visual stimulus onsets indicated with the black bars by a cutting window between -100 ms and 600 ms from each onset (out-of-window regions are colored gray). For the sake of visualization, the activities were smoothed with a Hann window of a 600 ms full width. The prototype (ii) has a specific temporal pattern of an initial strong activation in sparse codes (i.e., an inhibition in the alpha-band power) triggered by the onset cue. Following this initial event-related desynchronization, the alpha-band power moved slowly back to the baseline level, but relatively large residuals remained in the right condition even after 2 s from the Attention onset.

Figure 4.5 shows how the most discriminative dictionary atom, prototype (i) in Fig. 4.4, was modulated by subject-session-specific spatial transforms over different subjects and different sessions. Although they were from the same dictionary atom, they showed specific patterns for each subject; i.e., modulation of the prototype between sessions was weaker than that between subjects. This observation is consistent with the result shown in Fig. 4.4B.

4.3.2 Subject-transfer decoding

We evaluated the performance of the subject-transfer decoding. Before proceeding to the main analyses, we found that it was unable to decode the subjects' spatially attended direction from wavelet-transformed EOG signals under the one-session-out CV setting ($p > 0.05$, one-sided paired t-test, FDR corrected, avg. 51.5$\%$, s.d. 4.9$\%$). Thus, EOG artifacts did not contain enough information to allow us to decode spatial attention in our EEG measurements.

We compared our method with the following three baseline subject-transfer decoding methods:

(a) divCSP-MS: The divergence-based common spatial pattern (CSP) for multiple subjects proposed by Samek et al. (2014), which extracted a discriminative feature subspace independent of subjects.[*2]

(b) Ensemble: Sparse ensemble of classifiers derived from subject-session-specific temporal and spatial filters proposed by Fazli et al. (2009), where the ensemble was sparsified by ℓ_1 regularization to facilitate generalization for each target subject not included in the ensemble.[*3]

(c) BasicDL: The same subject-transfer settings with the proposed method but with dictionary learning based on the basic formulation (Eq. (4.2)).

Both (a) and (b) construct subject-invariant decoders that we can then attempt to generalize over target (unseen) subjects. Since we defined in this study that subject-transfer decoding may use resting-state data but may not use task data from target subjects, we did not evaluate here the performance by other subject-transfer decoding methods that need data when target subjects perform task sessions (Devlaminck et al., 2011; Kang and Choi, 2014; Lotte and Guan, 2010).

Figure 4.6 shows the subject-transfer decoding performance of the proposed method and the three baseline methods. The CV transfer performance by the proposed method was significantly higher than those by the baseline methods: (a) $p = 2.26 \times 10^{-7}$, (b) $p = 0.0291$, (c) $p = 0.0227$ with the Wilcoxon signed-rank test. The favorable performance by the proposed method compared with those by the existing methods, (a) and (b), can be attributed to the consideration of the intersubject variability in the proposed dictionary learning method.

There was no significant correlation between the transfer performance and the behavior performance defined by the accuracy of button pressing: (a) $p = 0.910$, (b) $p = 0.663$, (c) $p = 0.207$, (d) $p = 0.652$ with t-test.

[*2] We followed the settings by Samek et al. (2014): The frequency band was 8–30 Hz, the dimensionality of CSP patterns was six, and the divergence measure for extracting CSP was the symmetric beta Kullback-Leibler (KL) divergence whose parameter was selected from 0, 0.5, and 1 by a nested CV procedure. We used the implementation provided by Samek et al. (2014) for our analyses (*http://www.divergence-methods.org/*).

[*3] We modified the original method to use the same frequency bands with those in our method, whereas Fazli et al. (2009) used nine predefined bands with different widths and positions. The feature dimensionality used in CSP filters was four, which is the same as in the original study. The ℓ_1 regularization parameter was determined by nested CV.

To see which part of spatial activity patterns contributed to decoding of the spatial attention task, Fig. 4.7 shows frequency-wise dictionary atoms that had the highest absolute weights in the classifier. The spatial patterns estimated by the proposed method (upper row) were more focal than those by BasicDL (lower row). The prototypes obtained by the proposed method showed large activities around 1) right vIPS on the alpha-band ($2^{3.5} = 11.3$ Hz) and 2) right MT+ and the motor area on the low-gamma band ($2^5 = 32$ Hz). These regions have been found to be involved in the spatial attention task based on cortical currents estimation (Morioka et al., 2014; Siegel et al., 2008) and MEG analysis (Bauer et al., 2006). In contrast, the spatial patterns obtained by BasicDL were spread over different areas in wide frequency bands (Fig. 4.7, lower), making it difficult to observe which brain regions and frequencies were important for the decoding of spatial attention.

We examined the relationship between the subject-transfer decoding performance and the self-decoding performance by linear regression analysis.[*4] In the regression analysis, we drew a line $y = a(x-50)+b$, where the explanatory variable x was a self-decoding performance (in %) and the dependent variable y was a transfer performance (in %). We are interested in the intercept, or the *bias*, b because it represents the subject transfer performance for a virtual subject whose self-decoding performance is at the chance level; that is, by using data from other training subjects, we may be able to decode from the test subject whose self-decoding is difficult. By taking the dependent variable y to be one of (b) Ensemble, (c) BasicDL, or (d) ProposedDL, we performed least-square fitting for (b), (c), and (d) (Fig. 4.8). Data samples for the regression analysis were (x_i, y_i) $(i = 1, \dots, 41)$. The estimated intercepts and their p-values when compared to the chance level (50%) with t-test were (b) $b = 51.0$ ($p = 0.50$), (c) $b = 54.3$ ($p = 0.012$), and (d) $b = 54.7$ ($p = 0.0044$). Thus, the intercepts of (c) and (d) were significant.

[*4] The self-decoding performance denotes the decoding accuracy of a classifier trained using the target subject's data, which are not available in the subject-transfer scheme. For a self-decoding method, we employed CSP (Blankertz et al., 2008), one of the most popular methods for BMI. We followed the settings recommended in Blankertz et al. (2008): EEG signals were band-pass filtered in the 8–30 Hz frequency range, reduced to six dimensions using CSP filters, and then the log-variances of the six-dimensional data were used as features for linear discriminant analysis. For each target subject, the self-decoding performance was evaluated using a one-session-out CV procedure.

4.4 Discussion

In this study, we proposed a novel dictionary learning method that dissociates subject/session-independent spatial bases, subject-session-specific spatial transforms, and sparse codes, and designed a subject-transfer decoding scheme using resting-state brain activities as calibration data. Since this dissociation was effective in removing subject-session-specific variability in the EEG measurement signals, the obtained sparse codes provided robust features to subject-transfer decoding of the spatial attention task. Particularly in subject-transfer decoding, we estimated the subject-session specific transforms based on the resting-state data, with which we built decoders for target subjects without using calibration by task data. The transfer performance of our method was found to be superior to those of other subject-invariant decoding methods.

We suggest that the proposed dictionary learning method estimates dictionary atoms that are robust against the variability in the EEG measurements, which is supported by the consistency of the sparse codes over subjects and sessions (Fig. 4.4A). The consistency of the sparse codes implies that a limited set of dictionary atoms were frequently used in task sessions, so that the acquired dictionary was more focused for representing task-related brain activities. On the other hand, the basic dictionary learning method yielded sparse codes that were relatively inconsistent. The reason for this is that the dictionary estimated from the measurements involving variability had to prepare a variety of atoms (e.g., spatially shifted ones according to different sensor positions on the scalp), which resulted in low generalization capability. This consistency was indeed important for improving the decoding performance based on the sparse codes. Fig. 4.5 also supports the assumption that the spatial transforms modulated specific patterns for each subject-session. Accordingly, our dictionary learning satisfied our basic assumptions A1) and A2).

Since the spatial transform of the resting session was located close to those of the task sessions (Fig. 4.4B) and our subject-transfer decoding method using resting calibration performed favorably to the existing decoding methods, our approach of calibration using resting-state data based on the third assumption A3) has been justified. Since the physical configuration of EEG electrodes stays almost identical in a resting session and subsequent task sessions, the similarity of the transforms between these sessions can be naturally understood. Although recording resting-state activities is required by our subject-transform decoding scheme, it is much easier than conducting task sessions of several tens of minutes to calibrate the decoder; in resting sessions, subjects do not need to concentrate on a task but only need to stay quiet for several minutes. Previous neuroscience studies have reported that task-relevant evoked activities are embedded in a subspace of a larger space outlined by spontaneous activities (Kenet et al., 2003; Luczak et al., 2009); our re-use of the spatial transforms estimated from resting-state EEG activities could have been effective because of the extraction of such a subspace in a subject-specific manner. This unsupervised calibration method incorporating resting-state brain activities is the most notable feature of the proposed method, compared with the existing subject-invariant learning techniques.

The results in Fig. 4.4 indicate that the proposed method can also serve as a tool for reliable multi-subject-session analysis since the dictionary extracted essential brain activities relevant to the performed task. All of the brain regions activated in (i)–(iv) have been reported to be relevant to the selective spatial attention task (Siegel et al., 2008). In particular, we can observe hemispheric lateralization in the following two ways. First, prototypes (ii) and (iv) exhibited desynchronization in the posterior regions contralateral to the attended directions of right and left, respectively; this desynchronization in the alpha band has previously been demonstrated in a covert spatial attention task (Thut et al., 2006; Worden et al., 2000; Wyart and Baudry, 2008). Second, prototype (iii) involved alpha-synchronization in FEF contralateral to the attended direction. Moreover, the temporal activity patterns displayed in the right panels were also consistent with previous findings (Capilla et al., 2012; Grent-'t Jong et al., 2011; Händel et al., 2011).

Our study is a novel extension of the dictionary learning scheme to multisubject analyses under variability across subjects and sessions such to allow subject-transfer decoding with resting-calibration. Several studies have exploited dictionary learning for characterizing neural representation; e.g., spatio-temporal patterns of a P300 evoked potential (Barthélemy et al., 2013), spike sorting for electrophysiological data (Carlson et al., 2014), and resting state fMRI networks (Eavani et al., 2012). Some other studies have used dictionary learning and sparse coding for classifying EEG sensor signals (Hammer et al., 2011; Shin et al., 2012; Zhou et al., 2012), which attempt to decode subjects' intentions from sparse representations. Although Shin et al. (2012) proposed a decoding scheme of motor-imagery BMI based on sparse representation, they did not estimate the dictionary adaptively from data, but used a mere concatenation of training data after CSP filtering (Blankertz et al., 2008). Since these studies did not model variability across subjects and sessions explicitly, they lacked the ability of summarizing brain activities on the common signal representation domain and were not effectively applicable to multisubject data analysis. Kang and Choi (2014) have recently proposed to share a latent subspace of CSP among different subjects and tune it by subject-wise linear transforms for exploiting multisubject EEG data. They analyzed only labeled data during EEG recording of task sessions of target subjects. Our unsupervised feature extraction by dictionary learning is applicable to resting-state data, which achieves low-burden resting-calibration for subject-transfer decoding.

Another idea to compensate for the variability caused by different physical conditions in measurements, e.g., misalignment of EEG sensors and degradation of sensor impedance, would be to estimate cortical currents and to co-register them on a standard brain. However, cortical current estimation incurs an additional burden on practical applications because it requires the measurement of three-dimensional locations of EEG channels using a stylus and moreover structural MRI (Aihara et al., 2012; Morioka et al., 2014; Yoshimura et al., 2012). Cortical current estimation does not preclude the adoption of our dictionary learning technique, which can be applied to any kinds of signals including cortical currents and their derivatives. The combination of cortical current estimation and the compensation for the variability inherent to the brain activities by the proposed method is a promising approach to analyze multiple subject-session data in the cortical domain.

The good transfer-decoding performance (Fig. 4.6) and the focal weighted atoms (Fig. 4.7) imply that effective information sharing across subjects was achieved by the proposed method. Thanks to the signal representation in the common dictionary, we can compare activities from different subject-session pairs on the same standing point.

In the two-dimensional plane consisting of the self-decoding performance and the subject-transfer performance (Fig. 4.8), we found that the proposed method gained a significant decodability improvement especially when the self-decoding performance was in the chance level (50%). This result advocates that the proposed method is applicable not only to the subjects who showed favorable self-decodability but also those with poor self-decodability.

Since the poor self-decoding performance might be associated with non-discriminative EEG activities in the task sessions, subject-transfer decoding can be a good alternative. This is an advantage of the proposed method over the conventional subject-invariant decoding methods.

Although we presented in this study offline analysis results, the proposed method of subject-transfer decoding with resting-calibration can be applied to online experiments, where recorded brain signals are immediately decoded. This is because, once a common dictionary $\mathbf{D}$ and a target-subject-specific transform $\mathbf{Z}_{N0}$ are available, the feature vector supplied to the trained classifier, i.e., the sparse code for a single task trial of the target subject, can be computed rapidly (0.01 s with our Matlab with SPAMS implementation on a computer with Intel Xeon 3.6 GHz CPU and 96 GB memory). The calculation of the transform $\mathbf{Z}_{N0}$ took 97.5 s,[*5] which is affordable as a preprocessing time between the end of a resting session and the start of (possibly online) task sessions. The learning of the common dictionary $\mathbf{D}$ from the 40 training subjects' data took about 9 h, which has to be done before starting an experiment with the target subject.

It was unclear how many training subjects are sufficient to obtain a good transfer performance and what is the performance limit of the transfer decoding approach, and further investigations would be needed. In our preliminary experiments (Fig. 4.9),[*6] we observed that 1) at least five training subjects were required to achieve a significant performance increase compared to the chance level and 2) the transfer performance did not reach a plateau even with the full dataset consisting of 40 training subjects and one target (testing) subject. Using larger datasets may improve the transfer performance but there are many factors to be considered, including the characteristics (similarity) of subjects and the type of mental tasks.

For comparison, we also evaluated a task-calibration approach (Lotte and Guan, 2010) (Fig. 4.10) using our dataset and found that around 24–48 task trials of a target subject were necessary for the task-calibration method to realize a transfer performance comparable with that of the proposed resting-calibration method. The task-calibration method of Lotte and Guan

---

[*5] If $\mathbf{D}$ is fixed, $\mathbf{Z}_{N0}$ can be calculated only from the resting data of the target subject.

[*6] To evaluate the transfer performance in this additional experiment, we collected 41 values of transfer decoding accuracy for each N_{tr}; in the $N_{\text{tr}} = 40$ case we made full use of the 41-subjects dataset by employing one-subject-out CV, but when N_{tr} was less than 40 we performed a CV-like procedure as follows: For a given N_{tr}, we repeated training and testing 41 times, where in one repetition one target subject was used (differently between the repetitions) for testing and N_{tr} training subjects were randomly selected from the other 40 non-target subjects (when $N_{\text{tr}} = 40$ this procedure reduces to the one-subject-out CV).

(2010) estimates a CSP filter by combining a covariance matrix of target subject's task data and those of the other training subjects' task data.[*7] To see how the size of data from a target subject impacts on the decoding accuracy for the target subject, we carried out experiments to compute transfer performances with the number of task trials of the target subject (N_{trial}) equal to 12, 24, 36, 48, and 60, and found that the task-calibration method requires around 24–48 task trials of the target subject to be comparable with the proposed method. This result supports the advantage of the lower burden of the proposed method than the task-calibration method (Lotte and Guan, 2010) because 1) the proposed method needs only resting-state data and 2) the duration of the resting-state (5 minutes) is shorter than that of 24–48 task trials (about 8–16 minutes). Note that since N_{trial} was limited, these transfer performances were lower than the self-decoding accuracy.

There is some room for improving our methodology to make it applicable to a wider range of decoding and BMI. Temporal similarities of spatial transforms and sparse codes were not considered in the current formulation of dictionary learning (Eq. (4.4)), but incorporating them into the formulation will encourage the sparse codes to be more consistent over subjects and sessions, facilitating effective information sharing for subject-transfer decoding. This can be achieved by introducing additional constraints to the spatial transforms and/or sparse codes in the objective function in the dictionary learning. A similar modification can also be performed in the frequency domain; namely, we can apply a constraint encouraging close frequency bands to have similar properties. Spatial similarity of smoothness may also be introduced through regularizations or constraints to make spatially neighboring elements of $\mathbf{Z}_{ij}$ have similar values. Such incorporation of prior knowledge into the subject-session-specific transforms has a potential to enhance the performance of the proposed method. In the proposed dictionary learning scheme, no special constraint, other than the ℓ_2 norm constraint on each atom, was imposed on dictionary atoms. More problem-specific constraints, e.g., positivity constraints, on the dictionary may enhance its performance and interpretability. For this purpose adopting a variant of non-negative dictionary learning schemes (Dikmen and Févotte, 2011, 2012) would enhance

---

[*7] We used the implementation provided by Lotte and Guan (2011) for our analyses (*http://sites.google.com/site/fabienlotte/code-and-softwares/rcsptoolbox*). In this implementation, LDA is trained only from task data of the target subject. This strategy is slightly different from that proposed in Lotte and Guan (2010), in which they trained the LDA by combining data from both training subjects and a target subject.

the interpretability of the decomposed matrices. The automatic determination of the number of dictionary atoms is potentially possible based on, e.g., Bayesian nonparametrics (Zhou et al., 2009; Kang and Choi, 2014). Although we applied the proposed method only to the EEG signals, the methodology is applicable to other measurement modalities such as MEG and fMRI. Since the development of better classifiers was not the focus of this study, we used the linear SVM, as the most typical supervised classifier. Consideration of classifier design may be important for further improving the decoding performance and building a more widely applicable subject-transfer decoding method. Although the proposed method needs only resting resting-state data for a calibration, the online adaptation of the spatial bases and the classifier by using task data from the target subject is a promising approach to realize more reliable BMI system.

4.5 Summary

In this study, we proposed a novel dictionary learning scheme to extract spatial bases common across multiple subjects and sessions in an unsupervised manner and a new subject-transfer decoding method based on the features extracted via dictionary learning. To calibrate the decoder for a target subject, our subject-transfer decoding method used resting-state data from the target subject. Using resting-state activities for subject-transfer decoding is a novel concept and its relatively low burden for the target subject will enlarge the applicability of BMI decoding systems. Our newly developed method did not only outperform the existing subject-invariant decoding methods, but also extracted reasonable spatial bases common over multiple subjects by avoiding disturbance from subject- and session-specific variations. As we move toward the era of large-scale neuroscientific databases, our newly developed data-driven methodology can provide a new tool applicable to various studies in neuroengineering, such as the implementation of robust BMI systems working in realistic environments.

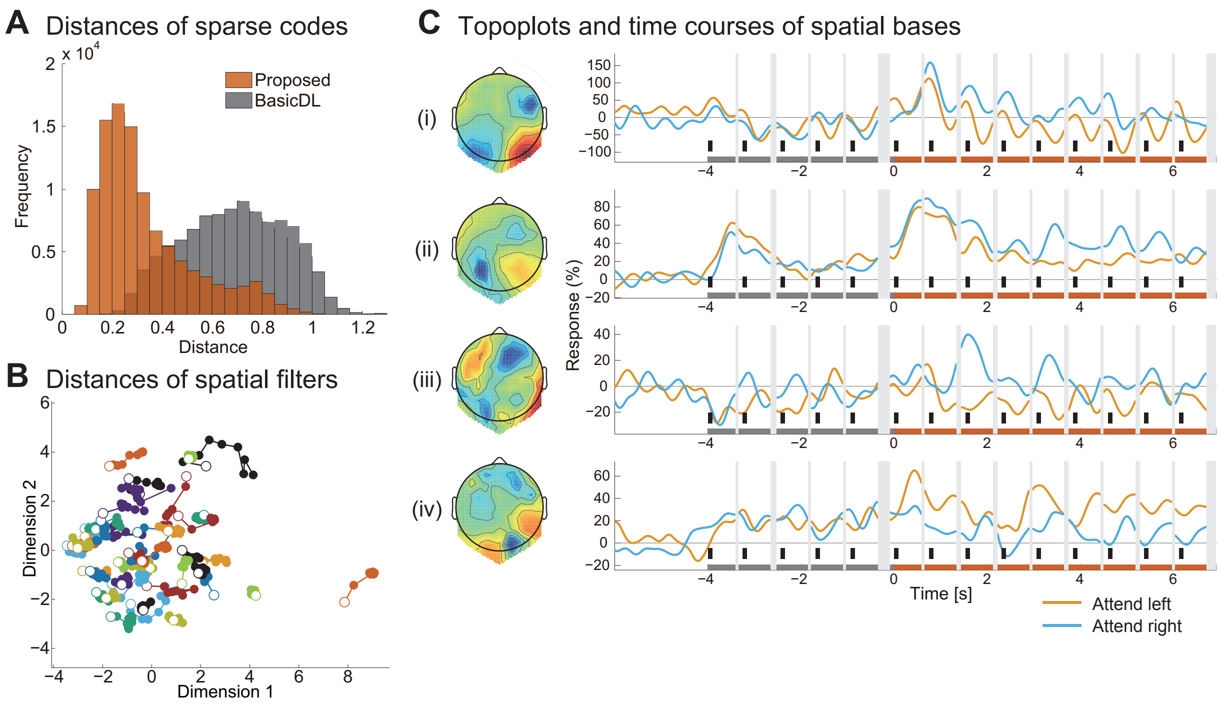


Fig. 4.4 Characteristics of the common spatial bases (dictionary), subject-session-specific transforms, and sparse codes, obtained by the proposed dictionary learning method. (A) Histograms of cosine distances of sparse codes between training and target subjects. The orange and gray histograms show the distances by the proposed formulation (Eq. (4.4)) and by the basic formulation (Eq. (4.2)), respectively. The introduction of the subject-session transform into the dictionary learning changed the distance distribution closer to the origin, implying that the sparse codes (to be used as decoding features) were made similar between the training and target subjects. (B) Multidimensional scaling of subject-session-specific spatial transforms $\mathbf{Z}_{ij}$. Each circle corresponds to a single session (indexed by j); the filled and open circles signify task and resting sessions, respectively. Temporally-adjacent sessions in the EEG experiment are connected by a line, and thus a single string of lines corresponds to one subject (indexed by i). The transforms of the same subject, including one for a resting session, are located nearby, implying they are consistent over resting and task sessions. (C) Prototypes (learned dictionary atoms) and the time courses of the associated sparse codes. The left panels show the prototypes with the top four largest differences (from (i) to (iv)) in the sparse codes between the two attention conditions, left and right. The right panels show the time courses of the sparse codes averaged over all the subjects for each task condition, “attend-left” (orange) and “attend-right” (blue). The horizontal axis in the right panels denotes the elapsed time relative to the onset of the Attention epoch. The vertical axis denotes the activity amplitude, in percent, relative to that in the Rest epoch (in a window between -8 s and -4 s from the Attention onset). The thick gray and red bars on the horizontal axis indicate the Control and Attention epochs, respectively.

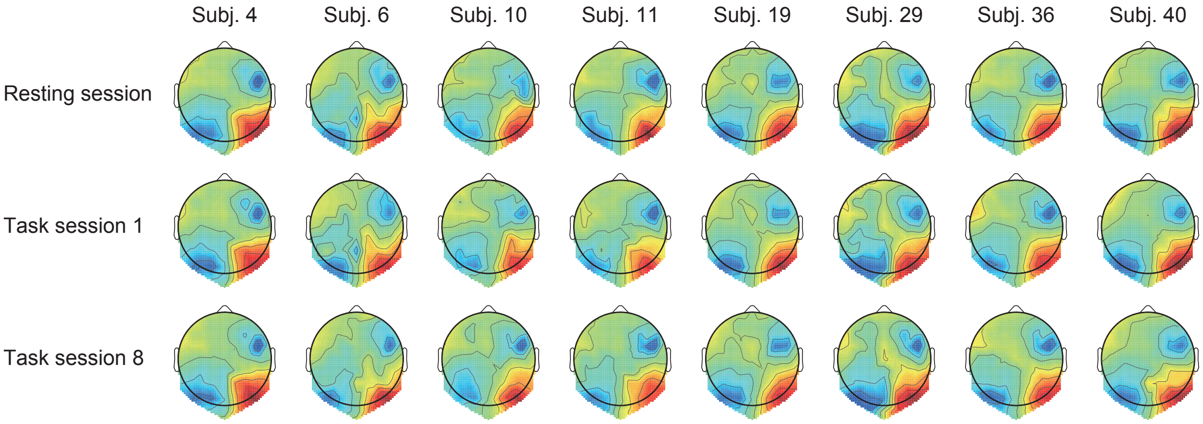


Fig. 4.5 Spatial patterns of the most discriminative atom, prototype (i) in Fig. 4.4, modulated by subject-session-specific spatial transforms, $\mathbf{Z}_{ij}\mathbf{d}_{(\mathrm{i})}$ for $i \in \{4, 6, 10, 11, 19, 29, 36, 40\}$ and $j \in \{0, 1, 8\}$. We show three sessions (resting (0th), first, and eighth sessions) of eight distinct subjects whose transforms were most distant from the center of mass in the MDS space (Fig. 4.4B).

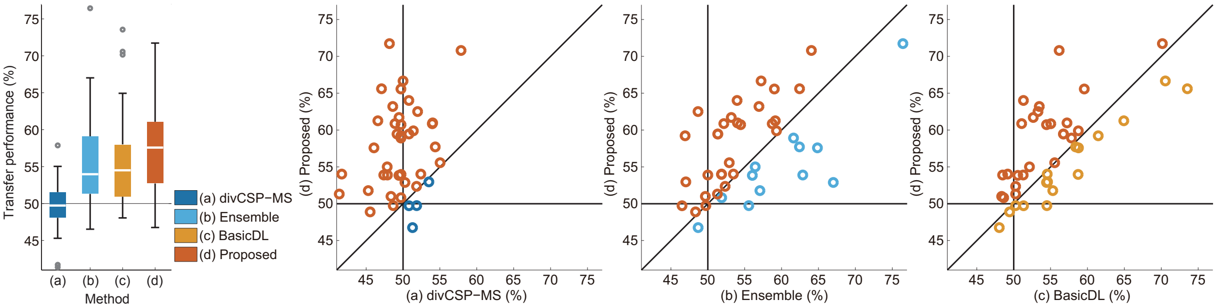


Fig. 4.6 Subject-transfer decoding performance by (a) divCSP-MS (Samek et al., 2014), (b) Ensemble (Fazli et al., 2009), (c) BasicDL (Eq. (4.2)), and (d) the proposed method. The leftmost panel shows boxplots of the transfer performances by the four methods, and the other panels show scatter plots comparing the transfer performances between the proposed method (d) and the three baseline methods (a−c). Each circle corresponds to a single target subject. The median transfer performance in the cross-validation (CV) procedure of the four methods was (a) 49.7%, (b) 54.0%, (c) 54.5%, and (d) 57.6%. The horizontal line shows the chance level (50%; not corrected for trial rejection).

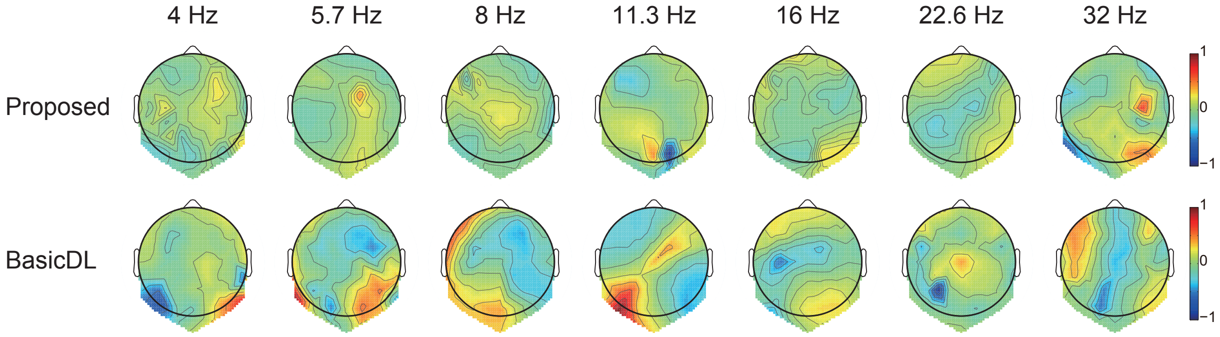


Fig. 4.7 Frequency-wise dictionary atoms with the largest absolute weights in the classifier. Upper: Dictionary atoms estimated by the proposed method. The dictionary and the classifier were trained using the training dataset in a CV repetition when the target subject was subject 38, who exhibited the highest transfer performance. Each column corresponds to a single frequency band. The dictionary atoms were normalized to make their values reside in the range $[-1, 1]$ after each was multiplied by the corresponding weight in the linear SVM classifier. Lower: Dictionary atoms estimated by BasicDL.

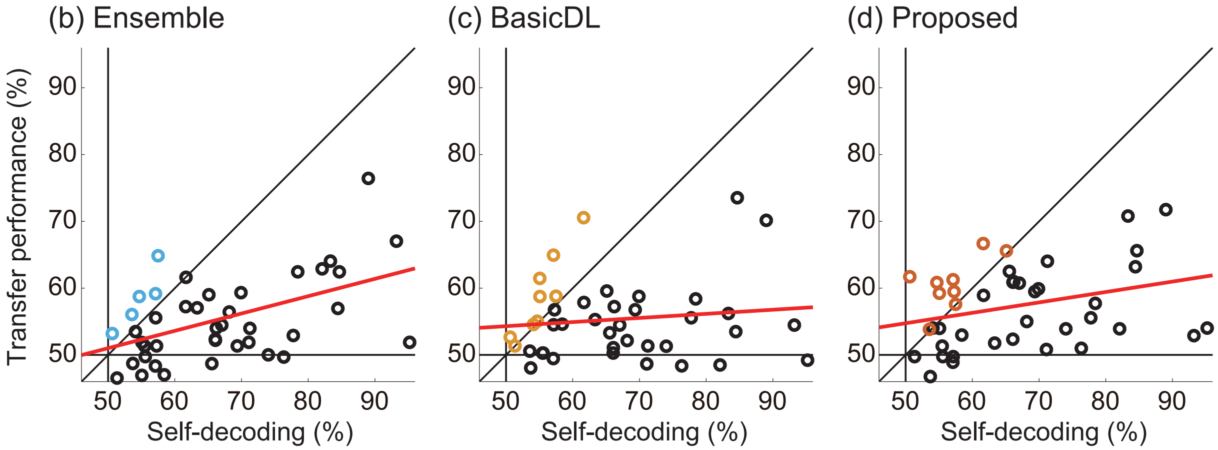


Fig. 4.8 Self-decoding performance (horizontal axis) vs. subject-transfer decoding performance (vertical axis); the latter was performed by (b) Ensemble (left panel), (c) BasicDL (middle panel), and (d) the proposed method (right panel). In each panel, a circle corresponds to a single target subject and is colored when the subject-transfer decoding performance was higher than the self-decoding performance (i.e., when the circle is plotted over the diagonal line, $y > x$). The colors for the over-the-diagonal circles are the same as in Fig. 4.6. Note that for each circle the x-value on the horizontal coordinate is common over the three panels, but its y-value on the vertical axis is different. The red lines are estimated least-square fits: (b) $y = 0.26(x - 50.0) + 51.0$, (c) $y = 0.062(x - 50.0) + 54.3$, and (d) $y = 0.16(x - 50.0) + 54.7$.

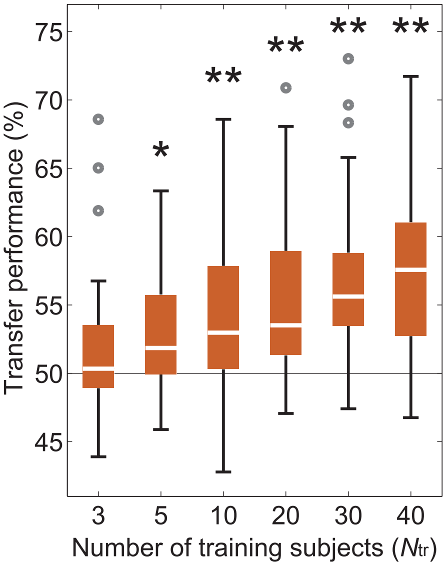


Fig. 4.9 The boxplots show the effect of the number of training subjects (N_{tr}, horizontal axis) to the transfer performance (vertical axis). Statistical significance in comparison to the chance level performance (50%) with the Wilcoxon signed-rank test: $*p < 0.001$; $**p < 0.0001$.

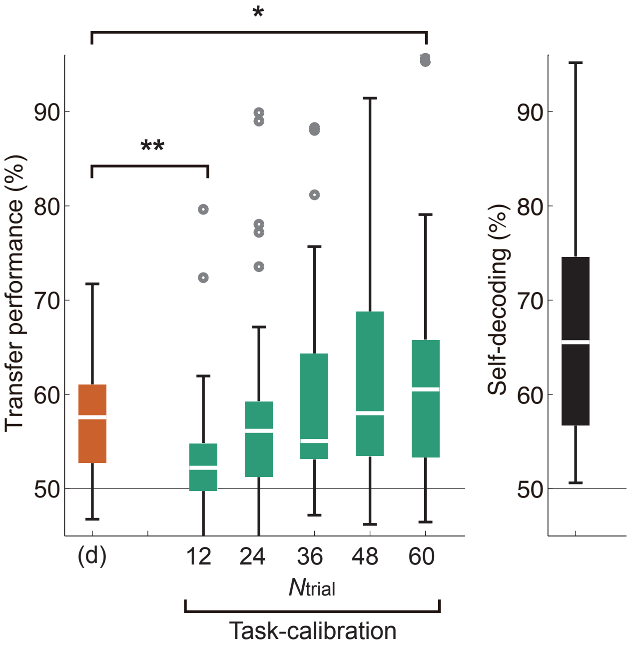


Fig. 4.10 (Left) The comparison with a task-calibration method (Lotte and Guan, 2010). The leftmost column shows the transfer performance of the proposed resting-calibration method (d) (same with Fig. 4.6), and the other columns show that of the task-calibration method with changing the number of task trials of a target subject (N_{trial}, horizontal axis). Statistical significance of comparison between the proposed method and the task-calibration method with the Wilcoxon signed-rank test: $*p < 0.05$; $**p < 0.01$. (Right) The boxplot shows the self-decoding performance same with Fig. 4.8.

Chapter 5

An autonomous navigation method toward BMI-controlled wheelchairs

5.1 Introduction

Robotic wheelchairs have been regarded as a major candidate for an application of BMI that improves the quality of life of handicapped people. Millions of people around the world suffer from mobility impairments, and it has been thought that between 1.4 and 2.1 million individuals might not be able to use a manual wheelchair because of lack of muscle control, and thus benefit from a smart powered wheelchair (Wästlund et al., 2010) that can provide autonomous assistances to the user based on the user's possible interfaces (e.g., joystick, voice, gaze, tongue, and bite) (Simpson et al., 2008). Some studies have proposed a wheelchair controlled by BMI in which the user can continuously give locomotion commands to the wheelchair (Galán et al., 2008; Hema et al., 2011). These BMI-wheelchairs can provide a novel methods for moving around to people who have severe motor-disabilities and are unable to use the conventional interfaces.

In recent years, to increase the reliability using BMI wheelchairs in daily life, we set up a BMI house that monitors and supports the wheelchair and the user living there (Kanemura et al., 2013) (Fig. 5.1). The BMI house installations and the wheelchair are fully electrically powered and are controllable by the BMI user's intentions through the central server. As a means of output for mobility, we use a robotic wheelchair that can navigate autonomously between different locations in a living environment using a waypoint-based BMI framework. Conventional BMI controllers (Carlson and Millan, 2013; Galán et al., 2008; Hema et al., 2011) continuously read the user's intentions from their brain activity throughout the operation period. This continuous control requires the user to constantly concentrate, which makes the user tired, and needs a high ITR (or decoding accuracy) of BMI.[*1] However, in general, such high self-decoding accuracies are limited to small group of subjects, and thus it is not always suitable for daily life support without a practical design. In the framework of waypoint-based BMI, to increase user comfort, the user of the wheelchair does not need to concentrate or control it continuously, but just needs to make decisions at an a priori waypoint prepared in the house to decide their subsequent destination (i.e., the next waypoint). This framework has the advantage that the user does not necessarily have to have a high ITR of BMI.[*2] The waypoints are placed on landmark spots that exist in the living environment. They are selected manually by the user at places such as in front of the entrance door, living room, bedroom, kitchen, and in front of television. At each waypoint there is a list of possible destinations from which the user can choose for the next direction the BMI-controlled wheelchair should go. For the decision making, they used Left/Right hand's motor imagery because it can be naturally associated with the direction of movement (Left/Right) and does not require an extra monitor for destination

---

[*1] It should be noted that in the asynchronous BMIs, the calculation of ITR is difficult because the timing of operation may vary dramatically depending on the user, and the user may choose not to send any commands for long periods. Thus, asynchronous BMIs often report decoding accuracies and task performances without using an ITR at all. For example, Carlson and Millan (2013) reported that four participants of their experiments had average self-decoding accuracies of 93.3%, 90.0%, 96.7%, and 100.0% with 2-class motor imagery task (Left hand/Right hand), and then successfully completed driving tasks twice without any collision; four participants in Hema et al. (2011) had average self-decoding accuracies of 96.5%, 95.7%, 87.7%, and 96.2% with 4-class motor imagery task (Relax/Both hands/Left hand/Right hand), and reached target points with a range of performances from 83.3% to 100%.

[*2] The participant in Kanemura et al. (2013) had self-decoding performance of 78.1%, which was lower than that reported in the asynchronous BMIs (Carlson and Millan, 2013; Hema et al., 2011) and comparable with that reported in our off-line experiment in Chapter 3 (79.1%), but the participant could perform BMI navigation with high success ratio of 91.7%.

Fig. 5.1 Experimental environment used in Kanemura et al. (2013). (Left) BMI house facade. (Right) Wheelchair in the living room.

selection.

Although semi-autonomous navigation reduces the user's stress and the load by not having to control the wheelchair continuously with a high ITR of BMI, the navigation framework needs robust navigation methods to stably and autonomously control the robotic wheelchair. There are two main modalities used to measure the environment for the robotic wheelchair's localization and navigation: the laser range finder (LRF) and the camera. In the system devised by Kanemura et al. (2013), the wheelchair localizes itself with a particle filter that corrects the position of the robot based on LRF scans using a laser data pre-processing method. Recently, vision-based mobile robots' SLAM and navigation have been the source of countless research contributions because of the rich sensory output and cost-effectiveness of the vision sensors. Many of these vision-based methods use local features such as corner points, scale-invariant feature transform (SIFT) (Lowe, 1999), and speeded-up robust features (SURF) (Bay et al., 2006) as landmarks. However, these systems cannot perform stable SLAM and navigation in crowded, bustling environments such as train stations and shopping malls, because when extracting feature points in such complicated environments, dynamic objects such as humans may disturb stable feature point detection. If such feature points are used as landmarks, the algorithm collapses and errors occur in map building and self-localization. These difficulties complicate the realization of semi-autonomous BMI-wheelchairs that can be used with human beings in real environments.

In this study, we propose a vision-based SLAM and navigation method that is stable even in crowded, bustling environments. In our method, we extract relatively stable feature points called position-invariant robust features (PIRF) (Kawewong et al., 2010, 2011) from monocular omnidirectional sequential images. PIRFs are extracted from sequential images by matching local feature points between all adjacent images and removing the feature points that do not match between certain adjacent images. Thus, we can eliminate numerous feature points that appear to be sensitive to changes in the observer's position. Because a PIRF is an image feature point, it cannot be directly applied to SLAM. Therefore, in this study, we propose a method to map PIRFs in three-dimensional (3D) space by combining local features extracted from sequential images and odometry information. In the proposed method, the mobile robot performs a learning phase and a navigation phase. In the learning phase, the mobile robot is controlled by a navigator and learns the route. In the navigation phase, the mobile robot plans the path on the learned route and moves to an arbitrary position autonomously. By using the proposed method, we can perform the SLAM and the navigation even in a crowded environment.

This chapter is organized as follows. Section 5.2 describes related works on mobile robot navigation in dynamic environments. The proposed method is described in Section 5.3. The experimental results are presented in Section 5.4, and we see that the mobile robot learns the route and moves autonomously in a crowded environment. Section 5.5 features concluding comments.

5.2 Related Work

Many conventional methods of SLAM and navigation in dynamic environments use LRFs. These methods propose the separation of LRF data into static and dynamic objects. Using this approach, they allow the application of LRF to SLAM and navigation in a dynamic environment. An example of a technique used for navigation in dynamic environments is Foka and Trahanias (2010). In this approach, the robot predicts the movement of people from the LRF

Fig. 5.2 The left image is an example of a crowded environment. This image was obtained by the robot in our experiment. Many people are constantly moving. The right image is the mobile robot used in this study. This robot has an omnidirectional camera at the top.

data so as to move autonomously, avoiding collisions with people; however, the robot had to be given a map in advance. There are also navigation methods in which the robot learns the map by itself in advance (Hatao et al., 2009; Müller et al., 2008). In Müller et al. (2008), map building had to be performed in a static environment, so there were difficulties in this method's practical application. In Hatao et al. (2009), the robot was able to build the map in a dynamic environment. However, the experiments were performed in a relatively low-population environment, so the method was not sufficiently robust.

In contrast to LRFs, the attention paid to vision-based SLAM and navigation methods has been growing in recent years. However, using a camera for SLAM and navigation in dynamic environments is more difficult than using LRFs. When extracting feature points from an image in a crowded environment, many feature points are extracted from not only static objects but also dynamic objects such as humans, making it difficult to separate static and dynamic objects. By recognizing such feature points as landmarks, the algorithm collapses and errors occur in map building and self-localization (Ballantyne et al., 2008).

Some examples of vision-based SLAM and navigation methods in dynamic environments are Booij et al. (2007); Gamallo et al. (2008); Koch et al. (2010); Shen and Hu (2006); Thrun et al. (1999). In Gamallo et al. (2008), the robot recognizes overhead lighting as landmarks using an omnidirectional camera that is pointing to the ceiling. Therefore, by reducing the influence of environmental features other than the overhead lighting, these methods can perform SLAM in a dynamic environment. However, these methods can be used only in certain environments in which there are overhead lights, and where data can be associated with these lights accurately. MINERVA (Thrun et al., 1999) is a museum guide robot, and has a camera pointing to the ceiling, similar to Gamallo et al. (2008). By combining the data from the camera with the LRF data, the robot can navigate in a crowded environment such as a museum. MINERVA can build a map by itself, but map-building has to be performed in a static environment. ATLAS (Shen and Hu, 2006) is a vision-based museum guide robot. In this situation, ATLAS made a vision-based topological map and localized itself using image-based matching algorithms. However, the topological map was too sparse to localize the robot with a high degree of accuracy; hence, ATLAS used LRF to localize itself more accurately. Booij et al. (2007); Koch et al. (2010) created the dictionary of local features by extracting SIFT features from images and built a topological map. These methods are similar to the proposed method; however, the types of feature points are different. In addition, these methods cannot learn a map in crowded environments, so they have to perform their initial map building in a static environment. Moreover, they build topological maps, so they cannot perform a shortest-path search from distance information.

In this study, we propose a vision-based method that performs map building and navigates to the given goal point from the given start point in crowded environments. In the proposed method, we perform robust SLAM and navigation in crowded environments by extracting stable feature points from sequential images and odometry, and building the map using these feature points. Using the proposed method, we can perform SLAM and navigation in more crowded environments. Here, the proposed method is based on a hybrid approach (Blanco et al., 2008; Bosse et al., 2004), which is a technique combining a metrical approach and topological approach. In the hybrid approach, each node contains local metric information, which we can use to perform shortest-path planning. For these reasons, we built a hybrid map and performed SLAM and navigation using the map.

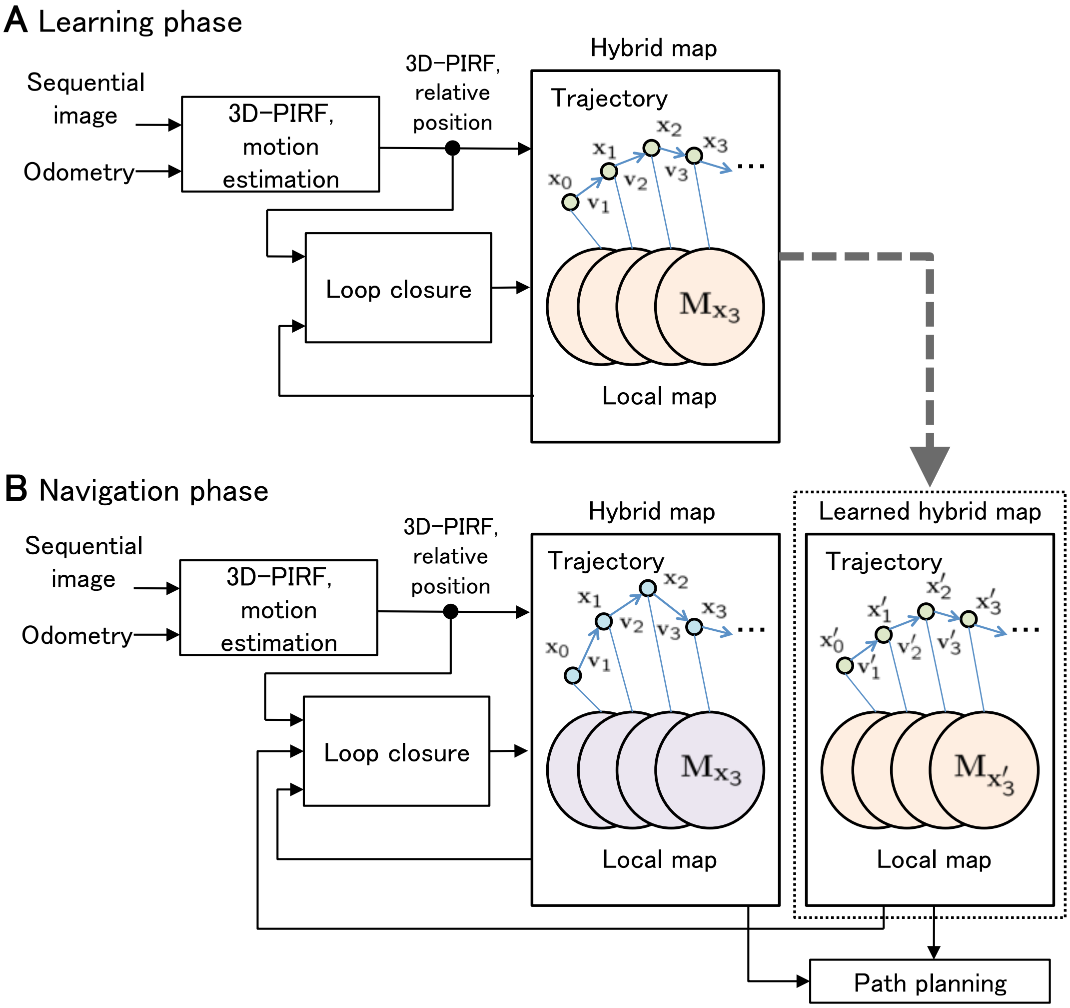


Fig. 5.3 Proposed system for SLAM and Navigation. (A) In the learning phase, the mobile robot is controlled by a navigator and learns trajectories to follow. By extracting three-dimensional position-invariant robust features (3D-PIRFs) from sequential images and odometry, the robot constructs a hybrid map. (B) In the navigation phase, the robot autonomously performs localization, path planning, and navigation on the basis of the learned trajectory.

5.3 Method

In this section, we propose a robust SLAM and navigation method by extracting 3D feature points from images that are stable even in crowded environments. We use a monocular omnidirectional camera and wheel encoder as sensors (Fig. 5.2). The proposed method is shown in Fig. 5.3. The following subsections describe the details of each process.

5.3.1 Stable 3D feature points extraction even in a crowded environment

Many vision-based methods use local feature points such as corner points, SIFT, and SURF. In general, it is easy to use local feature points to perform data association, so they are suitable for visual SLAM. However, when we extract these feature points from an image in a crowded environment, many feature points are extracted from dynamic objects and less-observable objects that become unobservable as the robot shifts a little. These unstable feature points impair map building and self-localization.

Recently, a new local feature point called the PIRF was proposed (Kawewong et al., 2010, 2011). The PIRF is a local feature point extracted from multiple sequential images by matching local feature points between all adjacent images and removing feature points that are not matched between certain adjacent images (Fig. 5.4). The number of sequential images (τ) for extracting the PIRF is set by the user. To extract PIRFs, first local feature points (e.g., corner point, SIFT, and SURF) are extracted for each image, and then matching of these feature points is performed sequentially for every adjacent pair of the past τ images, i.e., $(I_{t-\tau+1}, I_{t-\tau+2}), ..., (I_{t-1}, I_t)$. By tracing the matching information from the oldest image $I_{t-\tau+1}$ to the newest image I_t, feature points that appear repeatedly in all sequential images are retrieved; such feature points are called PIRFs (see Kawewong et al. (2010) for more details). The PIRF method can extract feature points for which the variation of appearance is small relative to the robot motion and, in contrast, eliminate feature points extracted from dynamic objects for which the variation of appearance is relatively large. Despite their simplicity, PIRFs are robust, especially against highly dynamic changes in scenes (Kawewong et al., 2010, 2011).

However, a PIRF is an image feature point, so it cannot be directly applied to hybrid SLAM. Thus, in this study, we propose a method to map PIRFs in 3D space by combining image feature points extracted from sequential images and odometry information. In this study, we call the proposed feature points mapped on 3D space 3D-PIRFs. Figure 5.5 outlines the 3D-PIRF concept. A 3D-PIRF is extracted as a 3D feature point by combining sequential images and odometry data obtained by the robot at each time step.

Here, because a 3D-PIRF is a 3D feature point, we have to estimate the 3D position of the feature point. Because the 3D-PIRF requires multiple observation steps, we require more sophisticated methods to estimate the 3D position. In general, because extended Kalman filter (EKF) methods incrementally record the observed feature points, the calculation time required to update the map increases quadratically. Accordingly, we used a formulation of a sliding window filter (SWF) (Sibley et al., 2007). A SWF is an intermediate method (between online methods such as EKF and offline methods) that reduces computational complexity by marginalizing old poses and feature points. This attribute is suitable for our method because, in the navigation phase, we do not need the old feature points (This is detailed in section 5.3.4). Using this method, we can estimate the 3D position of the 3D-PIRF.

Fig. 5.5 shows the SWF notations, where $\mathbf{X}_{t-\tau:t} = \{\mathbf{x}_{t-\tau}, ..., \mathbf{x}_t\}$ is the set of robot poses, τ is the length of the time window, and $\mathbf{M}_{t-\tau:t} = \{\mathbf{m}_1, ..., \mathbf{m}_n\}$ is the set of the feature point positions, which were observed in all steps from $t-\tau$ to t, $\mathbf{z}_{i,j}$ is the measurement of the i^{th} feature point observed from the j^{th} pose, and $\mathbf{u}_i$ is a control vector that was applied to the robot at time $i-1$ to take it to time i. Note that the image feature points are same as those extracted by the original PIRF (Fig. 5.4). The SWF method is used to estimate the maximum a posteriori (MAP) robot poses $\hat{\mathbf{X}}_{t-\tau:t}$ and the feature point positions $\hat{\mathbf{M}}_{t-\tau:t}$ from a set of odometry data $\mathbf{U}_{1:t} = \{\mathbf{u}_1, ..., \mathbf{u}_t\}$ and a set of measurements $\mathbf{Z}_{1:t} = \{\mathbf{z}_{1,1}, ..., \mathbf{z}_{n,t}\}$ by maximizing the posteriori $p(\mathbf{X}_{t-\tau:t}, \mathbf{M}_{t-\tau:t}|\mathbf{U}_{1:t}, \mathbf{Z}_{1:t})$ (by minimizing the negative log of the joint distribution $-\log(p(\mathbf{X}_{t-\tau:t}, \mathbf{M}_{t-\tau:t}, \mathbf{U}_{1:t}, \mathbf{Z}_{1:t})))$. Here we note that because 3D-PIRFs are extracted as feature points that are observed in each step, all 3D-PIRF feature points are observable from all robot poses from $t-\tau$ to t.

Consider the Bayesian network in Fig. 5.5, where each node represents a random variable in the system. The gray nodes represent observed variables, the white nodes represent hidden variables, and the arrows in the graph represent the dependence between variables.

The joint probability of $\mathbf{X}_{t-\tau:t}$, $\mathbf{M}_{t-\tau:t}$, $\mathbf{U}_{1:t}$, and $\mathbf{Z}_{1:t}$ can be factorized as follows:

$$\begin{aligned} p(\mathbf{X}_{t-\tau:t}, \mathbf{M}_{t-\tau:t}, \mathbf{U}_{1:t}, \mathbf{Z}_{1:t}) \propto & \; p(\mathbf{x}_{t-\tau}, \mathbf{M}_{t-\tau:t}|\mathbf{U}_{1:t-\tau}, \mathbf{Z}_{1:t-\tau}) \\ & \times \prod_{i=t-\tau+1}^{t} p(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{u}_i) \prod_j p(\mathbf{z}_j|\mathbf{x}_i, \mathbf{m}_j). \end{aligned} \tag{5.1}$$

Here $p(\mathbf{x}_{t-\tau}, \mathbf{M}_{t-\tau:t}|\mathbf{U}_{1:t-\tau}, \mathbf{Z}_{1:t-\tau})$ is the prior used at time t and is just the posterior at time $t-\tau$; that is, the distribution of vehicle pose and map at the beginning of the sliding window.

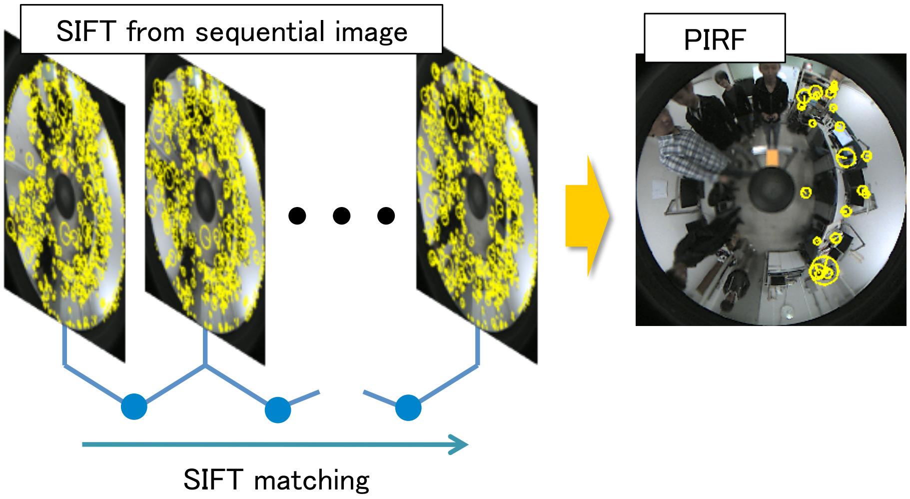


Fig. 5.4 PIRFs extracted from sequential images. The PIRF is a local feature extracted by matching local feature points of sequential images. The PIRF method can exclude local features showing motion.

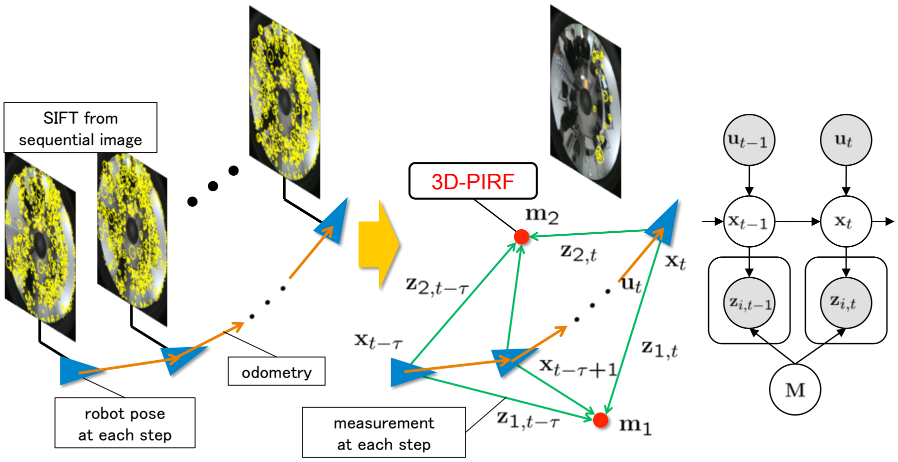


Fig. 5.5 3D-PIRFs extracted from sequential images and odometry. The 3D-PIRF is an extension of the PIRF and represents appearance and 3D position. In this figure, we show the simultaneous localization and mapping (SLAM) notations and a Bayesian network representation. Note that all feature points are observable at every step.

$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t)$ is the model of $\mathbf{x}_t$ (i.e., the probability of the new pose $\mathbf{x}_t$ given the last robot pose $\mathbf{x}_{t-1}$ and the odometry $\mathbf{u}_t$), and $p(\mathbf{z}_i|\mathbf{x}_t, \mathbf{m}_i)$ is the model of $\mathbf{z}_i$ (i.e., the probability of the measurement $\mathbf{z}_i$ given the robot pose $\mathbf{x}_t$ and feature point $\mathbf{m}_i$). Here, we make Gaussian assumptions as follows:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{u}_t) \propto \exp(-\frac{1}{2} \parallel f(\mathbf{x}_{t-1},\mathbf{u}_t)-\mathbf{x}_t \parallel^2_{Q_t}) \tag{5.2}$$

$$p(\mathbf{z}_i|\mathbf{x}_t,\mathbf{m}_i) \propto \exp(-\frac{1}{2} \parallel h(\mathbf{x}_t,\mathbf{m}_i) - \mathbf{z}_i \parallel^2_{R_t}). \tag{5.3}$$

Here, Q_t and R_t are covariances, f is the motion model, and h is the measurement model.

In experiments, we assume the motion model as

$$f(\mathbf{x}_{t-1},\mathbf{u}_t) = \mathbf{x}_{t-1} \oplus \mathbf{u}_t. \tag{5.4}$$

Here $\oplus$ denotes the transformation composition operator. We restrict the robot motion to the two-dimensional (2D) region; therefore, $\mathbf{x}_t = [x_{t,x}, x_{t,y}, x_{t,\theta}]^T$ is the position and angle of the robot in the 2D region, and $\mathbf{u}_t = [u_{t,x}, u_{t,y}, u_{t,\theta}]^T$ is the vector motion relative to pose $\mathbf{x}_{t-1}$ calculated by odometry. In this study, because we use a monocular camera, the measurement $\mathbf{z} = [z_h, z_v]^T$ is the feature point direction measured from the robot: z_h is the horizontal angle and z_v is the vertical angle. Thus, if the 3D feature position is assumed as $\mathbf{m}_i = [m_{i,x}, m_{i,y}, m_{i,z}]^T$, the measurement model h is

$$h(\mathbf{x}_t,\mathbf{m}_i) = \begin{bmatrix} \tan^{-1}\frac{m_{i,y}-x_{t,y}}{m_{i,x}-x_{t,x}} - x_{t,\theta} \\ \tan^{-1}\frac{m_{i,z}-x_{t,z}}{\sqrt{(m_{i,x}-x_{t,x})^2+(m_{i,y}-x_{t,y})^2}} \end{bmatrix} \tag{5.5}$$

which is the prediction of the measurement calculated from the robot pose and feature point position.

Using the SWF formulation, we can estimate the positions of 3D-PIRFs from multiple observation steps.

5.3.2 Hybrid map building by 3D-PIRF

In the proposed method, we use a hybrid approach, which facilitates path planning using the graph structure and shortest-path planning using the distance information. In Fig. 5.3, we show the hybrid map of the proposed method. The robot trajectory is represented by nodes and edges in the graph: each node represents local map information, and each edge represents the relative pose information $\mathbf{v}$ between the nodes.

First, we describe how to build the local map. We can estimate the mean and covariance of a 3D-PIRF position by using the method mentioned in the previous section. Thus, we can build the local map containing the 3D-PIRF information by transforming the mean and covariance to relative coordinates and preserving them.

Second, the edge information is obtained by calculating the relative pose between the nodes. We already estimated the robot poses at each time step, so using this result, the relative pose $\mathbf{v}_t$ is calculated from $\mathbf{x}_t \ominus \mathbf{x}_{t-1}$. Here $\ominus$ is the inverse motion composition operator; hence, $\mathbf{x}_t \ominus \mathbf{x}_{t-1}$ is the pose $\mathbf{x}_t$ relative to the pose $\mathbf{x}_{t-1}$.

5.3.3 Trajectory modification by loop closure

In general SLAM problems, the longer the distance the robot travels, the more errors accumulate. In the proposed method, the error is accumulated as a distortion of the trajectory. To correct the distortion of the trajectory, we need to recalculate the trajectory by using the information that the robot returned to a previously visited location (loop closure). In the proposed method, we use the previously proposed online loop closure detection method (Tongprasit et al., 2011). This method uses a PIRF as an image feature point and can detect loop closure online and incrementally in a crowded environment.

Fig. 5.6 shows the concept of correcting the trajectory by loop closure. The robot checks loop closure at each step. For example, we assume that the robot detects a loop closure between the current pose $\mathbf{x}_T$ and old pose $\mathbf{x}_0$. Here, for the convenience of using Eq. (5.6), we label $\mathbf{x}_T$ and $\mathbf{x}_0$ as $\mathbf{x}_{L(i)}$ and $\mathbf{x}_{L_0(i)}$, respectively. Next, we look for feature points matched between these local maps.

Then, we calculate the relative pose between these local maps by using the feature matching information between these local maps. In the proposed method, the relative pose α_i is calculated by an eight-point algorithm (Hartley and Zisserman, 2003). However, there is no guarantee that all feature point associations calculated by matching the feature point descriptors are correct, so we eliminate outliers by random sample consensus (RANSAC) (Fischler and Bolles, 1981).

Using the above results, we modify the trajectory $\{\mathbf{x}_0, ..., \mathbf{x}_N\}$ by minimizing the following expression:

$$\sum_{i=1}^{N_L} \| \mathbf{x}_{L_0(i)} \oplus \alpha_i - \mathbf{x}_{L(i)} \|_{P_L}^2 + \sum_{i=1}^{N} \| \mathbf{x}_{t-1} \oplus \mathbf{v}_t - \mathbf{x}_t \|_{P_\mathbf{v}}^2, \tag{5.6}$$

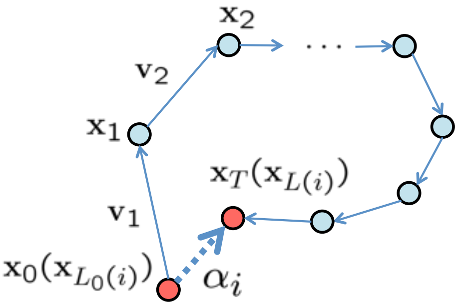


Fig. 5.6 Modifying the trajectory by loop closure. In step T, the robot detects a loop closure between $\mathbf{x}_T$ and $\mathbf{x}_0$. When the loop closure is detected, the robot estimates the relative pose α between these local maps by feature point correspondence and modifies the trajectory by minimizing Eq. (5.6).

where P_L and $P_{\mathbf{v}}$ are weights of the Mahalanobis distance, $\mathbf{x}_{L(i)}$ is the robot pose where the robot detected the loop closure with $\mathbf{x}_{L_0(i)}$, N_L is the number of loop closures, N is the total number of time steps, and $\mathbf{v}_t$ is the relative pose between adjacent nodes. By minimizing this expression, we can modify the trajectory.

In addition, when a loop closure is detected, we insert a new edge joined to the previously visited node in the hybrid map. Using this process, we can perform shortest path planning in the navigation phase.

5.3.4 Navigation

The goal of this study is mobile robot navigation, so the robot performs a learning phase and a navigation phase.

In the learning phase, the mobile robot is controlled by a navigator and learns the route by building the hybrid map using the method mentioned above.

Next, in the navigation phase, the robot navigates autonomously along the learned route. First, we specify the target point and start point that corresponds to the current robot pose. Thus, the robot searches the shortest path from the start point to the target point. In the proposed method, because we use a hybrid approach, the learned route is represented as a graph structure; thus, we can apply Dijkstra's algorithm (Dijkstra, 1959) to search for the shortest path on the graph.

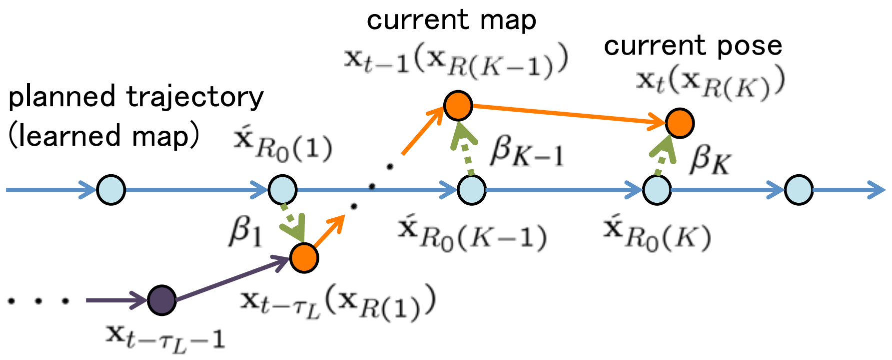


Fig. 5.7 Modifying the trajectory and estimating the current robot's pose by loop closure. The robot checks loop closure at each step in the learned hybrid map. If the robot detects a loop closure, it calculates the relative pose β between these local maps, modifies the trajectory, and estimates the current robot pose by minimizing Eq. (5.7).

Second, the robot navigates autonomously, building the hybrid map. In the navigation phase, the robot detects a loop closure as in the learning phase. However, in the navigation phase, the robot detects loop closures in the learned hybrid map. By detecting loop closures in the learned map and modifying the current trajectory, the robot can estimate its pose relative to the learned trajectory.

Fig. 5.7 illustrates the concept of estimating the robot pose in the navigation phase. The robot checks the loop closure at each step in the learned hybrid map. If the robot detects a loop closure between the current pose and the learned hybrid map, it calculates the relative pose β between these local maps using the same method mentioned in the previous section. Using the above results, we modify the trajectory $\{\mathbf{x}_{t-\tau_L}, ..., \mathbf{x}_t\}$ by minimizing the following expression:

$$\sum_{i\in S_R} \left\| \acute{\mathbf{x}}_{R_0(i)} \oplus \beta_i - \mathbf{x}_{R(i)} \right\|^2_{P_R} + \sum_{i=t-\tau_L}^{t} \| \mathbf{x}_{i-1} \oplus \mathbf{v}_i - \mathbf{x}_i \|^2_{P_{\mathbf{v}}}, \tag{5.7}$$

where $\acute{\mathbf{X}} = \{\acute{\mathbf{x}}_0, ..., \acute{\mathbf{x}}_N\}$ is the learned trajectory, which is fixed in the navigation phase; τ_L is the length of time window; P_R is the weight of the Mahalanobis distance; $\mathbf{x}_{R(i)}$ is the robot pose where the robot detected the loop closure with $\acute{\mathbf{x}}_{R_0(i)}$, which is in the learned map, and $S_R = \{i \mid t - \tau_L \leqq R(i) \leqq t$, is where the robot detected the loop closure in $\mathbf{x}_{R(i)}$ with the learned map.$\}$. By minimizing this expression, we can modify the trajectory. Here, in the navigation phase, we modify the trajectory by using the new information in the τ_L step, as well as using the SWF method, and we do not modify the old trajectory. Therefore, we can keep the calculation time for modifying the trajectory constant. By modifying the trajectory, the current robot pose $\mathbf{x}_t$ is also modified. Thus, we can estimate the robot pose using the loop closure information.

After the pose estimation, the robot moves toward the goal. Initially, the robot finds the closest node on the planned path and decides that the node several steps forward from this closest node is the subgoal. Next, the robot determines the velocity of the right and left wheels required to move toward the subgoal. Using these processes, the robot can autonomously navigate along the planned path and reach the goal using the shortest route. In our method, a robot can move along the edges of a learned hybrid map. The robot's direction and sequence of movement in the navigation phase are not limited to those in the learning phase. Hence, if the robot detects obstacles in a path, it can avoid the path or replan another path to reach the goal.

These are the processes of the proposed method. By considering this architecture, the robot can perform SLAM and navigation even in crowded environments.

5.4 Results and discussion

For the experiments, a iWs09 robot (Fig. 5.2) was used. The robot is equipped with an omnidirectional vision system consisting of a hyperbolic mirror and an ordinary camera with a resolution of 800 × 600 at a height of 1300 mm. In addition, the robot is equipped with encoder-equipped motors, which help detect the rotation of the motor. It also has a simple obstacle avoidance function using a 2D LRF to avoid collision with humans or obstacles. Note that we do not use the LRF to estimate the robot's position. The algorithms were created using Matlab and ran on an Intel Core2Duo computer (2.54 GHz, 2 GB) in the robot.

Feature point extraction in a dynamic environment

In this experiment, we showed that we could estimate the position of a 3D-PIRF.

The sliding window size τ was set at 4. The number of feature points depends on the robot's speed and the window size, and was set by the user. The larger the window size, the fewer the number of feature points. In contrast, when the window size is smaller, there are more feature points, so we cannot eliminate the dynamic objects. We extracted a 3D-PIRF from SURF, which can be calculated faster than SIFT and is rotation-invariant similar to SIFT; this attribute is suitable for an omnidirectional camera. We used standard feature matching, the same as that used by Lowe (2004). Here, the position estimation of the feature points in the front-back direction of the robot suffered from calculation errors because the measurement change was small relative to the robot's motion. Therefore, we eliminated SURF points that were in the front-back direction of the robot within a range of 40 degrees at each step. With this setup, we can estimate the 3D position of 3D-PIRF.

This experiment was performed in an indoor room about 4 m × 7 m, and five people were moving around in the left corner of the room. Fig. 5.8(a) shows the SURF feature points extracted from one image; the number of extracted SURF points was 1097. We see from this figure that many SURF points were extracted from the people who were moving around the robot.

In contrast, Fig. 5.8(b), Fig. 5.8(c), and Fig. 5.8(d) show the 3D-PIRFs extracted by the method mentioned in Subsection 5.3.1; the number of extracted 3D-PIRFs was 22. Fig. 5.8(b) shows the 3D-PIRFs with the image. We see from this figure that no feature points were extracted from the left corner in which people were moving. This result was derived from the process by matching the SURF between all adjacent images and removing the feature points that did not match between certain adjacent images. This process eliminated the feature points for which the variation of appearance was larger relative to the robot movement (e.g., moving people) and extracted the feature points for which the variation of appearance was small (e.g., desks, displays, and bookshelves). Fig. 5.8(c) and Fig. 5.8(d) show the plan view and elevation view of the estimated position of 3D-PIRFs, respectively. In these figures, circles indicate the 3D-PIRF position and triangles indicate the camera position. Here, the number of camera positions displayed in the figure was 4 because we set the sliding window size τ as 4. The position of a 3D-PIRF was estimated from the measurements obtained by the robot in these four places. We see from these figures that 3D-PIRF points were extracted from static objects such as desks, displays, and bookshelves, and distance errors between estimated positions of 3D-PIRFs and their ground truth positions were smaller than 200 mm. Fig. 5.8(e) shows the evaluation of the estimation of 3D-PIRF position. We evaluated the accuracy of the 3D-PIRF position by comparing the estimated and actual distance of the robot to the feature points. The estimated distance was calculated from the estimated current robot position and feature point positions, and the actual distance was measured manually. Here the combined number of the estimated and actual distances of each feature point was 4 because the sliding window size τ was 4; thus, we evaluated the error of a feature point by subtracting each estimated distance from the actual distance of the point and averaged the obtained values. The maximum error was 164 mm. From Fig. 5.8(e), we can see that the errors of the feature points that were closest to the robot (e. g., 10th–17th feature points) are relatively small. In contrast, the errors of the feature points that were furthest from the robot are relatively large.

As seen above, we can extract stable feature points even in crowded environments and estimate their position. Following the proposed method, we built a hybrid map containing extracted 3D-PIRF points, so that we could perform SLAM and navigation stably even in crowded environments.

Navigation in a crowded environment

In this subsection, we outline the navigation experiment and confirm the efficiency of the proposed method. This experiment was conducted in an indoor environment, a 20 m × 20 m sized school cafeteria. Fig. 5.2 is an image obtained by the robot in this experiment. In this environment, the general public was in disarray. In such a crowded environment, the conventional approaches could not perform SLAM and navigation stably because unstable features extracted from dynamic objects adversely affect map building and self-localization. Furthermore, in this environment, almost every wall surface was covered with glass, providing a huge amount of ambient light, making it somewhat difficult to obtain good feature points and perform SLAM and navigation. Nonetheless, the proposed method could perform SLAM and navigation stably even in such a crowded environment.

First, in the learning phase, the mobile robot was controlled by a navigator to learn a route by building the hybrid map. Fig. 5.9 shows a ground truth map and the route the robot adopted. At the beginning, we manually controlled the robot, moving it along the lower-left solid loop twice in the clockwise direction, then moving it to the upper right, and finally moving it along the large dotted loop twice in the clockwise direction. The total number of control steps was 1022, and in the learning phase, there were approximately 30–40 people in this environment. The maximum number of people concurrently observed by the omnidirectional camera was about 30. Note that because there was no ground truth of the robot's position, the route shown in the figure is not its true value.

Fig. 5.12(a) shows the trajectory calculated only from odometry. In the figure, the left circle is the start point, the upper-right circle is the goal point, and the calculated trajectory is described as a solid line. We see from the figure that the longer the distance the robot traveled, the more errors accumulated, so the robot could not learn the route correctly.

Fig. 5.10 shows an example of 3D-PIRFs extracted in this experiment. The number of extracted 3D-PIRFs was 44. We see from this figure that 3D-PIRFs were extracted from stably observable points (e.g., distant notices, fluorescent lights, and window frames), and were not extracted around moving people. Similarly, 3D-PIRFs were extracted only from stably observable objects, so we could perform SLAM and navigation stably even in this crowded environment.

Fig. 5.12(b) shows the trajectory learned by the proposed method. In the figure, red points indicate the places in which the robot detected a loop closure. We see from the figure that the trajectory was modified correctly even if the robot moved a long distance. However, we have no ground truth, so quantitative evaluation remains an issue. Here, the average number of 3D-PIRFs contained in a local map was 43.5.

In the next navigation phase, the robot navigated autonomously along the planned path using the process mentioned in Section 5.3.4. This time we designated a lower-left position as the start point, and a top-right position as the goal point. Thus, the robot planned the shortest path and navigated autonomously. In the navigation phase, there were approximately 60–70 people in this environment. The maximum number of people concurrently observed by the omnidirectional camera was about 30. Fig. 5.13 shows the robot trajectory in the navigation phase. In the figure, the gray line indicates the learned trajectory, the blue line indicates the robot trajectory in the navigation phase, and the circles indicate places in which the robot detected a loop closure. We see from this figure that by detecting a loop closure in the learned map and modifying the current trajectory, the robot could estimate its pose relative to the learned trajectory and navigate autonomously to the goal point along the learned trajectory. Fig. 5.11 shows the robot and environment in the navigation phase; the robot could navigate autonomously in a crowded environment. The total number of control steps was 172, the average number of

3D-PIRFs contained in a local map was 55.1, and the detection rate of loop closure was 44.2%.

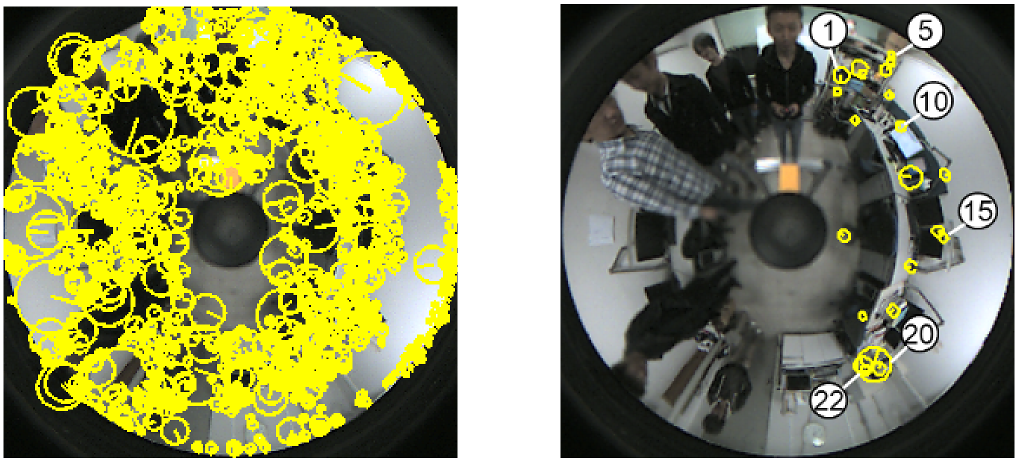
Here, the robot strayed from the planned path in a left-hand curve. This is because the robot could not detect the loop closure around this point and could not modify the localization error because many people were around the robot. The deviation from the planned route peaked here at 535 mm. The robot could make its way back to the planned route when the loop closure was detected again. The average deviation from the planned route was 150 mm. Here we could not observe the ground truth of the robot position; hence, we evaluated the deviation by comparing the learned hybrid map and the hybrid map in the navigation phase. The accuracy and robustness of the algorithm depended on the detection rate of loop closure. As long as the robot could detect a sufficient number of loop closures during τ_L of the previous steps, it could follow the learned path correctly, and the deviation from the learned map did not accumulate. Therefore, as τ_L increased, the accuracy and robustness of the algorithm increased. However, the calculation time also increased.

The average calculation time was 3.87 s and the average moving speed was 38.4 mm/s. So far, the average moving speed is relatively slow because the calculation time is large. In future work, we need to optimize the algorithm and rewrite the program in another language to accelerate the calculation. In addition, some feature fusion methods might be important for simplifying information.

5.5 Summary

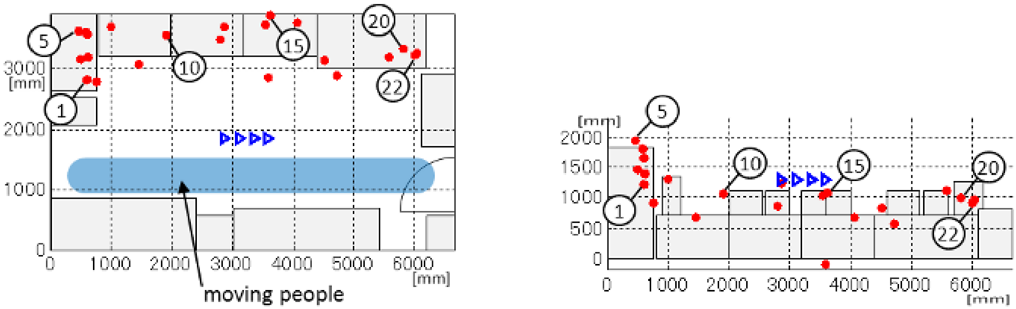
In this study, we proposed a vision-based SLAM and navigation method that is stable even in crowded environments. Conventional approaches cannot perform SLAM and navigation stably in crowded environments because unstable features are extracted from dynamic objects such as moving humans. In contrast, the proposed method produces stable SLAM and navigation even in crowded environments by extracting 3D-PIRF feature points that can eliminate feature points from dynamic objects. We confirmed the validity of the proposed method through experiments in crowded environments. First, we checked that 3D-PIRFs were extracted from stable points even in crowded environments. Second, we checked that the robot could learn a route without accumulated errors, and could navigate to a given goal point from a given start point along the learned route in the crowded environment.

This approach is necessary for robots to be able to adapt to the real world and interact with human beings, so we can expect it to contribute to the wide variety of practical applications for robots such as vehicles, humanoids, and BMI-controlled wheelchairs. The application of the proposed method to the BMI-controlled wheelchairs is a promising framework for supporting BMI users with low decoding accuracy and low ITR because the autonomous navigation technologies have the potential to complement BMI users' intentions with its autonomous systems.



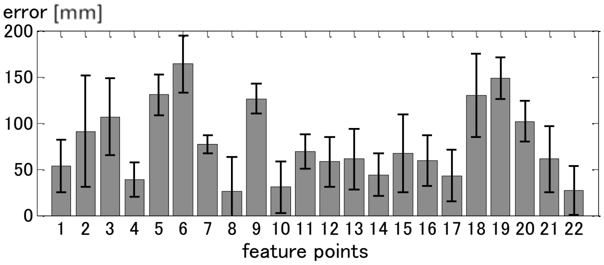
(a) SURF features. Many feature points were extracted from the places where people were moving.

(b) 3D-PIRF features. No feature points were extracted from the places where people were moving.



(c) Plan view of 3D-PIRF. The moving direction of the robot was right. Circles indicate feature positions, and triangles indicate robot poses. Colored area shows the area where people were moving.

(d) Elevation view of 3D-PIRF. The robot moved to the right. Circles indicate feature positions, and triangles indicate robot poses.



(e) Feature position errors evaluated with a ground truth.

Fig. 5.8 Results of experiment 1.

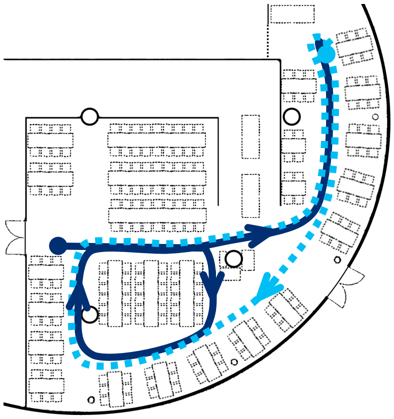


Fig. 5.9 The route in the learning phase. This environment is the same as in Fig. 5.2.

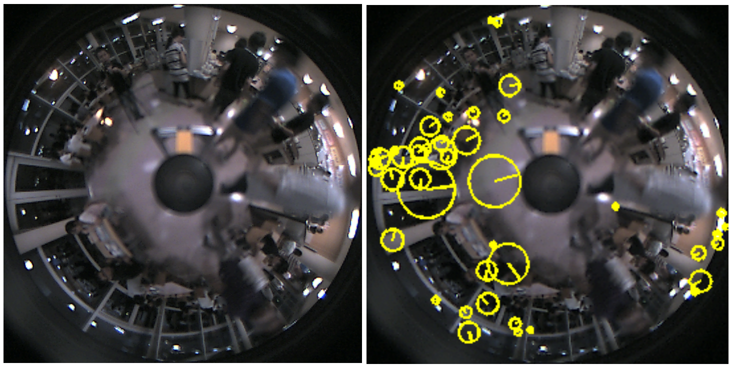
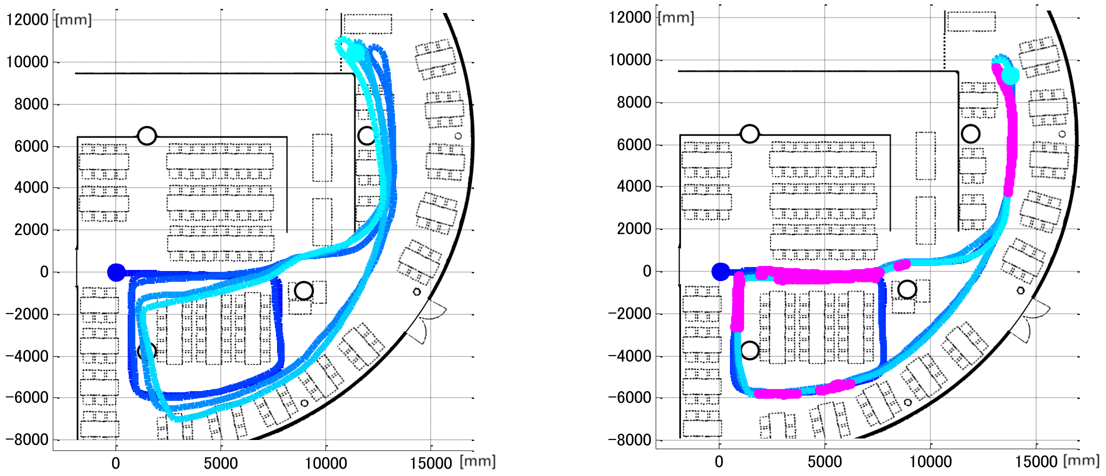


Fig. 5.10 3D-PIRFs. Feature points were extracted at the stable points.

Fig. 5.11 The robot and the environment in the navigation phase. The robot could navigate autonomously in a crowded environment.



(a) Trajectory calculated from only odometry. Errors accumulated gradually.

(b) Trajectory learned using the proposed method. The trajectory was modified from (a).

Fig. 5.12 Trajectory calculated from only odometry, and trajectory learned using the proposed method.

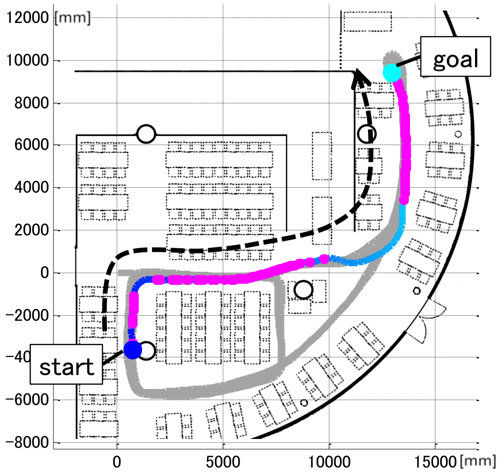


Fig. 5.13 Trajectory in the navigation phase. The robot was able to move autonomously from the start to the goal.

Chapter 6

Conclusions and Future Directions

In this thesis, novel methods for decoding from neuroimaging data and performing robot navigation, which can be applied to BMIs in real environments, have been presented.

The framework of cortical current estimation from EEG with NIRS prior information enables the analysis of human brain activity at the cortical level with high spatial resolution by avoiding the contamination of putative task-irrelevant activations. No previous studies have applied cortical current estimation from EEG with NIRS prior information to BMI decoding. We introduced a novel form of Bayesian prior to capture event related de-synchronization (i.e. modulations in frequency power) which is a well-known phenomenon during many tasks involving higher-order functions (e.g., spatial attention, and so on). By using sparsified logistic regression technique and then automatically selecting informative current sources for decoding, our new procedure outperformed major previous EEG decoding methods and another NIRS-EEG decoding method. Not only does our new procedure achieve significant accuracy improvements, but it also reveals reasonable activation patterns on the cortex, i.e. on IPS known to be involved in spatial attention, which allows us to check the validity of obtained neuro-decoders. Developments in this field potentially open up possibilities to investigate brain activities in naturalistic conditions, and hence are expected to develop powerful tools for BMI in real environments.

The second framework of the extraction of a dictionary common across subjects enables a robust analysis of multi-subject brain activity despite inherent variability across subjects. This methodology robustly extracted common spatial activity patterns that capture task-related ones while compensating for their inherent variability. Its application to subject-transfer decoding made it possible to employ resting-state brain activity to calibrate the decoder. This calibration technique is important for practical reasons because we can reduce the burden of conducting the standard task-performing calibration sessions for BMI users. We demonstrated the utility and validity of the methodology through its application to an EEG dataset of a selective spatial attention task, and showed that the method exhibited the best performance available. This study has a large impact as a novel methodology to robustly analyze multisubject brain activities under inherent variability across subjects, and as a novel concept in BMI to use resting-state brain activities for calibrating decoders specific to each user. The novelty would contribute to the neuroscience and neuroengineering communities and deserve broad interests because they will be important in the coming era of large-scale brain databases.

The third framework of a vision-based mobile robot's SLAM and navigation, which is workable even in crowded environments, is applicable to semi-autonomous BMI-controlled wheelchairs that coexist with human beings. Through experiments conducted in real and crowded environments, we showed that the proposed method was effective even in crowded environments by extracting robust 3D feature points from sequential vision images and odometry. The novelty would contribute to the practical application of BMI to wheelchairs that is workable in real environments, such as the BMI house where people are moving around.

These frameworks have the potential to be applied to BMIs in real environments and to enhance BMI performance, and hence are expected to impact on the community of real world neuroscience and neuroengineering.

Here we list remaining issues and possible directions that may be launched as immediate research topics.

For the first concept of cortical currents estimation (Chapter 3), there are several improvements in our methodology that are needed for applications to real environments. They include the reduction of the number of EEG electrodes and NIRS probes, and the efficient configuration of cortical current dipoles. The reduction of the number of EEG electrodes and NIRS probes is desirable for minimizing the physical burden on the subjects, and is also necessary for portable measurement systems since such systems' hardware would limit the affordable number of channels. According to Fig. 4.7 (left), which showed that the cortical currents effective for classification were located around fairly limited regions of the occipital and parietal areas, a strategy for minimizing the number of electrodes and probes could be considered. Aihara et al. (2012) showed that the combination of low-density EEG with NIRS-prior was better than high-density EEG without NIRS-prior. A natural question to ask next is how cortical current estimation behaves when the number of NIRS channels is decreased. Thus, further investigation on relationship between the number of NIRS channels and decoding accuracy is important for capturing more practical BMIs. The number of cortical currents is also an important factor. If we use a large number of cortical currents, we will suffer from the overfitting problem and a huge calculation cost. In contrast, if we use a small number of cortical currents, we cannot exploit the higher spatial resolution of cortical currents. Moreover, an appropriate number of cortical currents is related to the number of sensor signals. Thus, further investigation to determine the appropriate number of cortical currents is necessary for reliable BMIs in real environments. We used fixed ranges for band-pass filtering in the alpha and beta bands, but further improvements in decoding accuracy may be attained by optimizing the frequency band for each individual, because it is known that the spectral features of EEG signals have subject-dependent components (Ang et al., 2008).

For the second concept of learning a common dictionary (Chapter 4), there is some room for improving our methodology to make it applicable to a wider range of decoding and BMI. Temporal similarities of spatial transforms and sparse codes were not considered in the current formulation of dictionary learning (Eq. (4.4)), but incorporating them into the formulation will encourage the sparse codes to be more consistent over subjects and sessions, facilitating effective information sharing for subject-transfer decoding. This can be achieved by introducing additional constraints to the spatial transforms and/or sparse codes in the objective function in the dictionary learning. A similar modification can also be performed in the frequency domain; namely, we can introduce a constraint that encourages close frequency bands to have similar properties. In the proposed dictionary learning scheme, no special constraint, other than the ℓ_2 norm constraint on each atom, was imposed on the dictionary learning. More problem-specific constraints, e.g., positivity constraints, on the dictionary may enhance its performance and interpretability. The automatic determination of the number of dictionary atoms is potentially possible based on, e.g., Bayesian nonparametrics (Zhou et al., 2009). Although we applied the proposed method only to the EEG signals, the methodology is applicable to other measurement modalities such as MEG and fMRI. Since the development of better classifiers was not the focus of this study, we used the linear SVM, as the most typical supervised classifier. Consideration of classifier design may be important for further improving the decoding performance and building a more widely applicable subject-transfer decoding method. Although we extracted a common dictionary directly from EEG sensor signals, the common dictionary can also be extracted from cortical currents estimated in Chapter 3. The combination of compensation for physical variability by cortical current estimation and that for mental variability by learning a common dictionary is a promising approach to analyze multiple subject data in real environments in a more robust way. Whether to perform the cortical currents estimation depends on a tradeoff between data acquisition costs and a quality of signals.

Although these two methodologies can be applied to the analysis of neuro-imaging data in real environments, we evaluated their performance based only on datasets obtained in an experimental laboratory. Thus, the evaluation of the proposed methods using datasets obtained in more naturalistic conditions is necessary for further consideration of their applicability to BMIs in real environments.

For the third concept of a vision-based mobile robot's SLAM and navigation (Chapter 5), there are several improvements necessary for its practical applications. We see from the results that the calculation complexity was relatively high (about 3.87 s for each step), and then the robot's average moving speed was relatively slow (38.4 mm/s). Thus, for its practical application, we have to reduce the calculation complexity by optimizing the algorithm and its implementation. In this study, the robot only used an equipped camera for localizing itself. However, since the wheelchair being used in the BMI house is also equipped with LRFs, we would be able to increase the accuracy of position estimation by adding information obtained by the LRFs. Such a combination of information from the camera and that from the LRF is a promising approach to realize robust and stable BMI-controlled wheelchairs that coexist with human beings.

Although these three methodologies were proposed and evaluated separately, the combination of these methodologies is a promising approach to realize a BMI-controlled navigation system which is robust and comfortable for subjects. Currently, almost every system of the BMI-controlled navigation systems (e.g., Kanemura et al. (2013)) trains mental decoders specific to each subject directly from sensor signals. In general BMIs in real environments, such framework tends to suffer from external artifacts and contaminations by non-target brain activities, and imposes a lot of load on users for training the decoders specific to them. The combination of the proposed methods have advantages that we can decode the users' mental states without the contamination of sensor signals and we can save the users from burden of training the decoders and themselves. In addition, a combination of the newly-proposed vision-based robot navigation method and the conventional system (Kanemura et al., 2013) has a potential to provide a more robust complementary system to users who need external devices to move around the environment with BMI-based controllers.

Appendix A

VBMEG

Details of VBMEG are provided here and in previous reports (Sato et al., 2004; Yoshioka et al., 2008). The relationship between the amplitude of cortical currents and observed EEG sensor signals is described as

$$\mathbf{E} = \mathbf{G} \cdot \mathbf{J} + \boldsymbol{\epsilon}, \tag{A.1}$$

where $\mathbf{E} \in \mathbb{R}^{N\times 1}$ is the observed EEG signals, $\mathbf{J} = \begin{bmatrix}\mathbf{J}_{\text{brain}}^{\mathrm{T}} & \mathbf{J}_{\text{eye}}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ is a vector of current sources comprised of $\mathbf{J}_{\text{brain}} \in \mathbb{R}^{L\times 1}$, corresponding to cortical currents and $\mathbf{J}_{\text{eye}} \in \mathbb{R}^{K\times 1}$, corresponding to “extra” eye currents, and finally $\boldsymbol{\epsilon} \in \mathbb{R}^{N\times 1}$ is observation noise. Constants N, L, and K denote the number of EEG sensors, cortical current sources, and eye current sources, respectively. Here, $\mathbf{G} = \begin{bmatrix}\mathbf{G}_{\text{brain}} & \mathbf{G}_{\text{eye}}\end{bmatrix}$, where $\mathbf{G}_{\text{brain}} \in \mathbb{R}^{N\times L}$ and $\mathbf{G}_{\text{eye}} \in \mathbb{R}^{N\times K}$ are lead field matrices that describe the sensitivity of the sensors to the source currents. These lead fields are calculated from structural MRI data or its approximation by spherical models. Assuming that the EEG observation noise $\boldsymbol{\epsilon}$ obeys a Gaussian distribution with a spherical covariance $\beta^{-1}\mathbf{I}$, where β is an unknown inverse variance, we have the likelihood function

$$P(\mathbf{E}|\mathbf{J}) \propto \exp\left[-\frac{\beta}{2}\|\mathbf{E} - \mathbf{G}\cdot\mathbf{J}\|^2\right]. \tag{A.2}$$

The prior probability distribution of $\mathbf{J}$ is assumed to be

$$P_0(\mathbf{J}|\boldsymbol{\alpha}) \propto \exp\left[-\frac{1}{2}\mathbf{J}^{\mathrm{T}}\cdot\operatorname{diag}(\boldsymbol{\alpha})\cdot\mathbf{J}\right], \tag{A.3}$$

where $\boldsymbol{\alpha} = \begin{bmatrix}\boldsymbol{\alpha}_{\text{brain}}^{\mathrm{T}} & \boldsymbol{\alpha}_{\text{eye}}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, and $\boldsymbol{\alpha}_{\text{brain}} \in \mathbb{R}^{L\times 1}$ and $\boldsymbol{\alpha}_{\text{eye}} \in \mathbb{R}^{K\times 1}$ are the inverse variances of the brain and eye current sources, respectively. VBMEG estimates the current inverse variances

$\boldsymbol{\alpha}$ by introducing an ARD hierarchical prior as follows:

$$P_0(\boldsymbol{\alpha}) = \prod_i \Gamma(\alpha_i|\bar{\alpha}_{0i}, \gamma_0), \tag{A.4}$$

where i is a current source index running from 1 to $L + K$, $\Gamma(\alpha_i|\bar{\alpha}_{0i}, \gamma_0)$ is the Gamma distribution with mean $\bar{\alpha}_{0i}$ and degree of freedom γ_0.

The design of the hyperparameters in the hierarchical prior $\bar{\nu}_{0i} = \bar{\alpha}_{0i}^{-1}$ and γ_0 characterizes VBMEG estimates. If $\bar{\nu}_{0i}$ is large, prior knowledge implies that estimated currents should have large variances whereas small values of $\bar{\nu}_{0i}$ encourages estimated currents to be close to zero, leading to sparse solutions. The definition of the mean prior variance $\bar{\nu}_{0i}$, with hyperparameters m_0 and γ_0, has been discussed in Section 3.2.7.

Bibliography

Aihara, T., Takeda, Y., Takeda, K., Yasuda, W., Sato, T., Otaka, Y., Hanakawa, T., Honda, M., Liu, M., Kawato, M., Sato, M., Osu, R., 2012. Cortical current source estimation from electroencephalography in combination with near-infrared spectroscopy as a hierarchical prior. NeuroImage 59, 4006–4021.

Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., Wolpaw, J. R., 2008. Towards an independent brain–computer interface using steady state visual evoked potentials. Clin. Neurophysiol. 119 (2), 399 – 408.

Andersson, P., Pluim, J. P. W., Siero, J. C. W., Klein, S., Viergever, M. A., Ramsey, N. F., 2011. Real-time decoding of brain responses to visuospatial attention using 7T fMRI. PLoS ONE 6, 11.

Ang, K. K., Chin, Z. Y., Zhang, H., Guan, C., 2008. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. pp. 2390–2397.

Baldassarre, A., Lewis, C. M., Committeri, G., Snyder, A. Z., Romani, G. L., Corbetta, M., 2012. Individual variability in functional connectivity predicts performance of a perceptual task. Proc. Natl. Acad. Sci. 109 (9), 3516–3521.

Ballantyne, J., Valibeik, S., Darzi, A., Yang, G.-Z., 2008. Robotic navigation in crowded environments: Key challenges for autonomous navigation systems. In: Proc. 8th Workshop Perform. Metr. Intell. Syst. pp. 306–312.

Barthélemy, Q., Gouy-Pailler, C., Isaac, Y., Souloumiac, A., Larue, A., Mars, J., 2013. Multivariate temporal dictionary learning for EEG. J. Neurosci. Method. 215 (1), 19 – 28.

Bauer, M., Oostenveld, R., Peeters, M., Fries, P., 2006. Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. J. Neurosci. 26 (2), 490–501.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features. In: Comput. Vis. - ECCV. Vol. 3951. pp. 404–417.

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., Perelmouter, J., Taub, E., Flor, H., 1999. A spelling device for the paralysed. Nature 398 (6725), 297–298.

Birbaumer, N., Hinterberger, T., Kubler, A., Neumann, N., 2003. The thought-translation device (ttd): neurobehavioral mechanisms and clinical outcome. IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2), 120–123.

Blanco, J.-L., Fernandez-Madrigal, J.-A., Gonzalez, J., 2008. Toward a unified bayesian approach to hybrid metric–topological slam. IEEE Trans. Robot. 24 (2), 259–270.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Curio, G., 2007. The non-invasive berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects. NeuroImage 37 (2), 539–550.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process. Mag. 25 (1), 41–56.

Boas, D. A., Dale, A. M., Franceschini, M. A., 2004. Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. NeuroImage 23, Supplement 1 (0), S275–S288.

Booij, O., Terwijn, B., Zivkovic, Z., Krose, B., 2007. Navigation using an appearance based topological map. In: IEEE Int. Conf. Robot. Autom. pp. 3927–3932.

Bosse, M., Newman, P., Leonard, J., Teller, S., 2004. Simultaneous localization and map building in large-scale cyclic environments using the atlas framework. Int. J. Robot. Res. 23 (12), 1113–1139.

Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., Barnes, G. R., Smith, S. M., Morris, P. G., 2011. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. Proc. Natl. Acad. Sci. 108 (40), 16783–16788.

Brunner, C., Allison, B. Z., Krusienski, D. J., Kaiser, V., Müller-Putz, G. R., Pfurtscheller, G., Neuper, C., 2010. Improved signal processing approaches in an offline simulation of a hybrid brain–computer interface. J. Neurosci. Method 188 (1), 165 – 173.

Buch, E., Weber, C., Cohen, L. G., Braun, C., Dimyan, M. A., Ard, T., Mellinger, J., Caria, A., Soekadar, S., Fourkas, A., Birbaumer, N., 2008. Think to move: a neuromagnetic brain-computer interface (bci) system for chronic stroke. Stroke 39 (3), 910–917.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., Liu, T. T., 2004. Modeling the hemodynamic response to brain activation. NeuroImage 23, Supplement 1 (0), S220 – S233.

Capilla, A., Schoffelen, J.-M., Paterson, G., Thut, G., Gross, J., 2012. Dissociated α-band modulations in the dorsal and ventral visual pathways in visuospatial attention and perception. Cerebral Cortex.

Capotosto, P., Babiloni, C., Romani, G. L., Corbetta, M., 2009. Frontoparietal cortex controls spatial attention through modulation of anticipatory alpha rhythms. J. Neurosci. 29, 5863–5872.

Carlson, D., Vogelstein, J., Wu, Q., Lian, W., Zhou, M., Stoetzner, C., Kipke, D., Weber, D., Dunson, D., Carin, L., 2014. Multichannel electrophysiological spike sorting via joint dictionary learning and mixture modeling. IEEE Trans. Biomed. Eng. 61 (1), 41–54.

Carlson, T., Millan, J. d. R., 2013. Brain-controlled wheelchairs: A robotic architecture. IEEE Robot. Autom. Mag. 20 (1), 65–73.

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., Nicolelis, M. A. L., 2003. Learning to control a brain–machine interface for reaching and grasping by primates. PLoS Comput. Biol. 1 (2), e42.

Chan, A. M., Halgren, E., Marinkovic, K., Cash, S. S., 2011. Decoding word and category-specific spatiotemporal representations from MEG and EEG. NeuroImage 54, 3028–3039.

Chao, Z. C., Nagasaka, Y., Fujii, N., 2010. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. Frontiers in Neuroengineering 3 (3).

Chevallier, S., Barthelemy, Q., Atif, J., 2014. Subspace metrics for multivariate dictionaries and application to EEG. In: Int. Conf. Acoust. Speech Signal Process. (ICASSP) 2014. pp. 7178–7182.

Cohen, D., 1968. Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. Science 161 (3843), 784–786.

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J., Velliste, M., Boninger, M. L., Schwartz, A. B., 2013. High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet 381 (9866), 557–564.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., Shulman, G. L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3, 292–297.

Coyle, S. M., Ward, T. E., Markham, C. M., 2007. Brain-computer interface using a simplified functional near-infrared spectroscopy system. J. Neural. Eng. 4, 219.

Cuffin, B. N., Cohen, D., 1979. Comparison of the magnetoencephalogram and electroencephalogram. Electroencephalogr. Clin. Neurophysiol. 47 (2), 132 – 146.

Daitch, A. L., Sharma, M., Roland, J. L., Astafiev, S. V., Bundy, D. T., Gaona, C. M., Snyder, A. Z., Shulman, G. L., Leuthardt, E. C., Corbetta, M., 2013. Frequency-specific mechanism links human brain networks for spatial attention. Proc. Natl. Acad. Sci. 110 (48), 19585–19590.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., Halgren, E., 2000. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67.

de Pasquale, F., Penna, S. D., Snyder, A. Z., Marzetti, L., Pizzella, V., Romani, G. L., Corbetta, M., 2012. A cortical core for dynamic integration of functional networks in the resting human brain. Neuron 74 (4), 753–764.

Devlaminck, D., Wyns, B., Grosse-Wentrup, M., Otte, G., Santens, P., 2011. Multisubject learning for common spatial patterns in motor-imagery bci. Comput. Intell. Neurosci. 2011 (8).

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1 (1), 269–271.

Dikmen, O., Févotte, C., 2011. Maximum marginal likelihood estimation for nonnegative dictionary learning. In: IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP). pp. 1992–1995.

Dikmen, O., Févotte, C., 2012. Maximum marginal likelihood estimation for nonnegative dictionary learning in the gamma-poisson model. IEEE Trans. Signal Process. 60 (10), 5163–5175.

Do, A., Wang, P., King, C., Chun, S., Nenadic, Z., 2013. Brain-computer interface controlled robotic gait orthosis. J. NeuroEng. Rehabil. 10 (1).

Donchin, E., Coles, M. G. H., 1988. Is the p300 component a manifestation of context updating? Behav. Brain Sci. 11 (3), 357–374.

Dornhege, G., Millán, J. d. R., Hinterberger, T., McFarland, D. J., Müller, K.-R., 2007. Toward Brain-Computer Interfacing. MIT Press.

Doud, A. J., Lucas, J. P., Pisansky, M. T., He, B., 2011. Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. PLoS ONE 6 (10), e26322.

Eavani, H., Filipovych, R., Davatzikos, C., Satterthwaite, T., Gur, R., Gur, R., 2012. Sparse dictionary learning of resting state fmri networks. In: Int. Workshop Pattern Recognit. Neuroimaging (PRNI) 2012. pp. 73–76.

Elad, M., 2010. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer.

Falk, T. H., Guirgis, M., Power, S., Chau, T. T., 2011. Taking NIRS-BCIs outside the lab: Towards achieving robustness against environment noise. IEEE Trans. Neural System Rehabil. Eng. 19, 136–146.

Farwell, L., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. 70 (6), 510 – 523.

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Muller, K. R., Blankertz, B., 2012. Enhanced performance by a hybrid NIRS-EEG brain computer interface. NeuroImage 59, 519–529.

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.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Foka, A., Trahanias, P., 2010. Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. Int. J. Soc. Robot. 2 (1), 79–94.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., Raichle, M. E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. 102 (27), 9673–9678.

Friston, K. J., 2009. Modalities, modes, and models in functional neuroimaging. Science 326 (5951), 399–403.

Galán, F., Nuttin, M., Lew, E., Ferrez, P., Vanacker, G., Philips, J., Millán, J. d. R., 2008. A brain-actuated wheelchair: Asynchronous and non-invasive brain–computer interfaces for continuous control of robots. Clin. Neurophysiol. 119 (9), 2159–2169.

Gamallo, C., Regueiro, C., Quintia, P., Mucientes, M., 2008. Montecarlo localization for a guide mobile robot in a crowded environment based on omnivision. In: Towards Auton. Robot. Syst. pp. 1–8.

Gao, S., IW., T., Chia, L. T., Zhao, P., 2010. Local features are not lonely—laplacian sparse coding for image classification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). pp. 3555–3561.

Garrett, D. D., Kovacevic, N., McIntosh, A. R., Grady, C. L., 2011. The importance of being variable. J. Neurosci. 31 (12), 4496–4503.

Georgopoulos, A., Schwartz, A., Kettner, R., 1986. Neuronal population coding of movement direction. Science 233 (4771), 1416–1419.

Graimann, B., Allison, B., Pfurtscheller, G., 2011. Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction. Springer.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., Vanrumste, B., 2008. Review on solving the inverse problem in EEG source analysis. J. NeuroEng. Rehabil. 5 (1), 25.

Grent-'t Jong, T., Boehler, C. N., Kenemans, J. L., Woldorff, M. G., 2011. Differential functional roles of slow-wave and oscillatory-alpha activity in visual sensory cortex during anticipatory visual–spatial attention. Cerebral Cortex.

Grynszpan, F., Geselowitz, D. B., 1973. Model studies of the magnetocardiogram. Biophys. J. 13 (9), 911 – 925.

Güllmar, D., Haueisen, J., R., R. J., 2010. Influence of anisotropic electrical conductivity in white matter tissue on the EEG/MEG forward and inverse solution. a high-resolution whole head simulation study. NeuroImage 51 (1), 145–163.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., Lounasmaa, O. V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65 (2), 413–497.

Hammer, B., Chavarriage, R., Millán, J. d. R., 2011. Learning dictionaries of spatial and temporal EEG primitives for brain-computer interfaces. In: Workshop Structured Sparsity: Learning and Inference, ICML.

Händel, B. F., Haarmeier, T., Jensen, O., 2011. Alpha oscillations correlate with the successful inhibition of unattended stimuli. J. Cognit. Neurosci. 23 (9), 2494–2502.

Hartley, R., Zisserman, A., 2003. Multiple view geometry in computer vision, secon edition. Cambridge University Press.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition. Springer.

Hatao, N., Hanai, R., Yamazaki, K., Inaba, M., 2009. Real-time navigation for a personal mobility in an environment with pedestrians. In: IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN). pp. 619–626.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430.

Haynes, J. D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7, 523–534.

Hema, C., Paulraj, M., Yaacob, S., Adom, A., Nagarajan, R., 2011. Asynchronous brain machine interface-based control of a wheelchair. In: Arabnia, H. R., Tran, Q.-N. (Eds.), Software Tools and Algorithms for Biological Systems. Vol. 696. Springer New York, pp. 565–572.

Hermes, D., Miller, K. J., Vansteensel, M. J., Aarnoutse, E. J., Leijten, F. S., Ramsey, N. F., 2012. Neurophysiologic correlates of fmri in human motor cortex. Hum. Brain Mapp. 33 (7), 1689–1699.

Hinterberger, T., Kübler, A., Kaiser, J., Neumann, N., Birbaumer, N., 2003. A brain–computer interface (bci) for the locked-in: comparison of different {EEG} classifications for the thought translation device. Clin. Neurophysiol. 114 (3), 416–425.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Voge, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., Donoghue, J. P., 2012. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature 485, 372–375.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., Donoghue, J. P., 2006. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442, 164–171.

Hoffmann, U., Vesin, J. M., Ebrahimi, T., Diserens, K., 2008. An efficient P300-based brain-computer interface for disabled subjects. J. Neurosci. Method. 167, 115–125.

Holmes, G. L., 1999. Event-related desynchronization. handbook of electroencephalography and clinical neurophysiology, revised series, volume 6. Epilepsy Res. 49 (2).

Hopfinger, J. B., Buonocore, M. H., Mangun, G. R., 2000. The neural mechanisms of top-down attentional control. Nat. Neurosci. 3, 284–291.

Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y., 2013. Neural decoding of visual imagery during sleep. Science 340 (6132), 639–642.

Hoshi, Y., Kobayashi, N., Tamura, M., 2001. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. J. Appl. Physiol. 90, 1657–1662.

Hunt, J. J., Dayan, P., Goodhill, G. J., 2013. Sparse coding can predict primary visual cortex receptive field changes induced by abnormal visual input. PLoS Comput. Biol. 9 (5), e1003005.

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., Boas, D. A., 2006. A temporal comparison of bold, asl, and NIRS hemodynamic responses to motor stimuli in adult humans. NeuroImage 29 (2), 368 – 382.

Iturrate, I., Antelis, J. M., Kubler, A., Minguez, J., 2009. A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation. IEEE Trans. Robot. 25, 614–627.

Jobsis, F., 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198 (4323), 1264–1267.

Kajihara, S., Ohtani, Y., Goda, N., Tanigawa, M., Toyama, K., 2004. Wiener filter magnetoencephalography of visual cortical activities. Brain Topogr. 17, 13–25.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kanemura, A., Morales, Y., Kawanabe, M., Morioka, H., Kallakuri, N., Ikeda, T., Miyashita, T., Hagita, N., Ishii, S., 2013. A waypoint-based framework in brain-controlled smart home environments: Brain interfaces, domotics, and robotics integration. In: IEEE/RSJ Conf. Intell. Robt. Syst. (IROS). pp. 865–870.

Kang, H., Choi, S., 2014. Bayesian common spatial patterns for multi-subject EEG classification. Neural Netw. 57 (0), 39–50.

Kastner, S., Ungerleider, L. G., 2000. Mechanisms of visual attention in the human cortex.

Annu. Rev. Neurosci. 23, 315–341.

Kauhanen, L., Nykopp, T., Lehtonen, J., Jylanki, P., Heikkonen, J., Rantanen, P., Alaranta, H., Sams, M., 2006. Eeg and meg brain-computer interface for tetraplegic patients. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2), 190–193.

Kawewong, A., Tangruamsub, S., Hasegawa, O., 2010. Position-invariant robust features for long-term recognition of dynamic outdoor scenes. IEICE Trans. Inf. Syst. E93-D (9), 2587–2601.

Kawewong, A., Tongprasit, N., Tangruamsub, S., Hasegawa, O., 2011. Online and incremental appearance-based slam in highly dynamic environments. Int. J. Robot. Res. 30, 33–55.

Kelly, S. P., Lalor, E., Reilly, R. B., Foxe, J. J., 2005a. Independent brain computer interface control using visual spatial attention-dependent modulations of parieto-occipital alpha. In Conf. Proc. 2nd Int. IEEE EMBS Conf. on Neural Eng., 667–670.

Kelly, S. P., Lalor, E. C., Finucane, C., McDarby, G., Reilly, R. B., 2005b. Visual spatial attention control in an independent brain-computer interface. IEEE Trans. Biomed. Eng. 52, 1588–1596.

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., Arieli, A., 2003. Spontaneously emerging cortical representations of visual attributes. Nature 425, 954–956.

Koch, O., Walter, M., Huang, A., Teller, S., 2010. Ground robot navigation using uncalibrated cameras. In: IEEE Int. Conf. Robot. Autom. pp. 2423–2430.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., Baker, C. I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.

Kübler, A., Neumann, N., 2005. Brain-computer interfaces — the key for the conscious brain locked into a paralyzed body. Prog. Brain Res. 150 (513 - 525).

Lalor, E. C., Kelly, S. P., Finucane, C., Burke, R., Smith, R., Reilly, R. B., McDarby, G., 2005. Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment. EURASIP J. Appl. Signal Process. 2005, 3156–3164.

Leeb, R., Friedman, D., Muller-Putz, G. R., Scherer, R., Slater, M., 2007. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic. Comput. Intell. Neurosci. 2007, 79642.

Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G., Moran, D. W., 2004. A brain–computer interface using electrocorticographic signals in humans. J. Neural Eng. 1 (2), 63.

Logothetis, N. K., 2008. What we can do and what we cannot do with fmri. Nature 453 (7197), 869–878.

Lotte, F., Guan, C., 2010. Learning from other subjects helps reducing brain-computer interface calibration time. In: IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP). pp. 614–617.

Lotte, F., Guan, C., 2011. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. IEEE Trans. Biomed. Eng. 58 (2), 355–362.

Lowe, D., 1999. Object recognition from local scale-invariant features. In: IEEE Int. Conf. Comput. Vis. Vol. 2. pp. 1150–1157 vol.2.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Luczak, A., Barthó, P., Harris, K. D., 2009. Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron 62 (3), 413–425.

Ma, L., Zhang, L., Wang, L., Xu, M., Qi, H., Wan, B., Ming, D., Hu, Y., 2012. A hybrid brain-computer interface combining the eeg and nirs. IEEE Int. Conf. Virtual Environ. Hum.-Comput. Interfaces Meas. Systems (VECIMS) 2012, 159–162.

Mairal, J., Bach, F., Ponce, J., 2012. Task-driven dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. 34 (4), 791–804.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11, 19–60.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A., 2008. Supervised dictionary learning. In: Adv. Neural Inf. Process. Syst. Vol. 21.

Mallat, S., 2008. A Wavelet Tour of Signal Processing: The Sparse Way, 3rd Edition. Academic Press.

Mallat, S., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. 41 (12), 3397–3415.

Massar, S. A., Kenemans, J. L., Schutter, D. J., 2014. Resting-state EEG theta activity and risk learning: sensitivity to reward or punishment? Int. J. Psychophysiol. 91 (3), 172–177.

McFarland, D. J., Sarnacki, W. A., Wolpaw, J. R., 2010. Electroencephalographic (eeg) control of three-dimensional movement. J. Neural Eng. 7 (3), 036007.

McIntosh, A. R., Vakorin, V., Kovacevic, N., Wang, H., Diaconescu, A., Protzner, A. B., 2013. Spatiotemporal dependency of age-related changes in brain signal variability. Cereb. Cortex.

Mehring, C., Nawrot, M. P., de Oliveira, S. C., Vaadia, E., Schulze-Bonhage, A., Aertsen, A., Ball, T., 2004. Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex. J. Physiol.-Paris 98 (406), 498–506.

Mehring, C., Rickert, J., Vaadia, E., de Oliveira, S. C., Aertsen, A., Rotter, S., 2003. Inference of hand movements from local field potentials in monkey motor cortex. Nat. Neurosci. 6, 1253–1254.

Mellinger, J., Schalk, G., Braun, C., Preissl, H., Rosenstiel, W., Birbaumer, N., Kübler, A., 2007. An meg-based brain–computer interface (bci). NeuroImage 36 (3), 581 – 593.

Michel, C., Murray, M., Lantz, G., Gonzalez, S., Spineli, L., Grave de Peralta, R., 2004. EEG source imaging. Clin. Neurophysiol. 115, 2195–2222.

Middendorf, M., Mcmillan, G., Calhoun, G., Jones, K. S., 2000. Brain-computer interfaces based on the steady-state visual-evoked response. IEEE Trans. Rehabil. Eng. 8 (2), 211–214.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H., Sadato, N., Kamitani, Y., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60 (5), 915 – 929.

Morioka, H., Kanemura, A., Morimoto, S., Yoshioka, T., Oba, S., Kawanabe, M., Ishii, S., 2014. Decoding spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior information. NeuroImage 90 (15), 128–139.

Morishige, K., Kawawaki, D., Yoshioka, T., Sato, M., Kawato, M., 2009. Artifact removal using simultaneous current estimation of noise and cortical sources. In: Int. Conf. Neural Information Processing (ICONIP) 2008. Vol. 5506 of Lecture Notes in Computer Science. Springer, pp. 336–343.

Mosher, J., Leahy, R., Lewis, P., 1999. EEG and MEG: forward solutions for inverse methods. IEEE Trans. Biomed. Eng. 46 (3), 245–259.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 77 (3), 586–595.

Müller, J., Stachniss, C., Arras, K. O., Burgard, W., 2008. Socially inspired motion planning for mobile robots in populated environments. In: Int. Conf. Cognit. Syst.

Müller-Putz, G. R., Scherer, R., Brauneis, C., Pfurtscheller, G., 2005. Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components. J. Neural Eng. 2 (4), 123.

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., Andersen, R. A., 2004. Cognitive control signals for neural prosthetics. Science 305 (5681), 258–262.

Neuper, C., Müller-Putz, G. R., Scherer, R., Pfurtscheller, G., 2006. Motor imagery and eeg-based control of spelling devices and neuroprostheses. Prog. Brain Res. 159, 393–409.

Neuper, C., Scherer, R., Wriessnegger, S., Pfurtscheller, G., 2009. Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain–computer interface. Clin. Neurophysiol. 120 (2), 239–247.

Nijboer, F., Sellers, E., Mellinger, J., Jordan, M., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D., Vaughan, T., Wolpaw, J., Birbaumer, N., Kübler, A., 2008. A p300-based brain–computer interface for people with amyotrophic lateral sclerosis. Clin. Neurophysiol. 119 (8), 1909 – 1916.

Nirkko, A. C., C., O., Redmond, S. M., Bürki, M., Schroth, G., Hess, C. W., Wiesendanger, M., 2001. Different ipsilateral representations for distal and proximal movements in the sensorimotor cortex: Activation and deactivation patterns. NeuroImage 13 (5), 825–835.

O'Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., Nicolelis, M. A. L., 2011. Active tactile exploration using a brain-machine-brain interface. Nature 479, 228–231.

Ogawa, S., Lee, T.-M., Nayak, A. S., Glynn, P., 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn. Reson. Med. 14 (1), 68–78.

Olshausen, B. A., Field, D. J., 1997. Sparse coding with an overcomplete basis set: A strategy employed by v1? Vis. Res. 37 (23), 3311–3325,.

Pascual-Marqui, R. D., Michel, C. M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. 18, 49–65.

Patel, V. M., Chellappa, R., 2013. Sparse Representations and Compressive Sensing for Imaging and Vision. Springer.

Penny, W., Friston, K., Ashburner, J., Kiebel, S., Nichols, T., 2006. Statistical Parametric Mapping: The Analysis of Functional Brain Images, 1st Edition. Academic Press.

Pfurtscheller, G., Aranibar, A., 1979. Evaluation of event-related desynchronization (erd) preceding and following voluntary self-paced movement. Electroencephalogr. Clin. Neurophysiol. 46 (2), 138–146.

Pfurtscheller, G., Flotzinger, D., Kalcher, J., 1993. Brain-computer interface—a new communication device for handicapped persons. J. Microcomput. Appl. 16 (3), 293–299.

Pfurtscheller, G., Guger, C., Müller, G., Krausz, G., Neuper, C., 2000. Brain oscillations control hand orthosis in a tetraplegic. Neurosci. Lett. 292 (3), 211–214.

Pfurtscheller, G., Müller, G. R., Pfurtscheller, J., Gerner, H. J., Rupp, R., 2003. 'thought' – control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. Neurosci. Lett. 351 (1), 33–36.

Pfurtscheller, G., Neuper, C., 2006. Future prospects of erd/ers in the context of brain–computer interface (bci) developments. In: Neuper, C., Klimesch, W. (Eds.), Event-Related Dynamics of Brain Oscillations. Vol. 159 of Prog. Brain Res. Elsevier, pp. 433 – 437.

Pfurtscheller, G., Solis-Escalante, T., Ortner, R., Linortner, P., Muller-Putz, G., 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 (4), 409–414.

Phillips, C., Rugg, M. D., Friston, K. J., 2002. Anatomically informed basis functions for EEG source localization: Combining functional and anatomical constraints. NeuroImage 16, 678–695.

Power, S. D., Kushki, A., Chau, T., 2011. Towards a system-paced near-infrared spectroscopy brain–computer interface: differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state. J. Neural Eng. 8 (6), 066004.

Ramsey, N., van de Heuvel, M., Kho, K., Leijten, F., 2006. Towards human bci applications based on cognitive brain systems: an investigation of neural signals recorded from the dorsolateral prefrontal cortex. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2), 214–217.

Rebsamen, B., Guan, C., Zhang, H., Wang, C., Teo, C., Ang, M. H., Burdet, E., 2010. A brain controlled wheelchair to navigate in familiar environments. IEEE Trans. Neural Syst. Rehabil. Eng. 18, 590–598.

Samek, W., Kawanabe, M., Muller, K.-R., 2014. Divergence-based framework for common spatial patterns algorithms. IEEE Rev. Biomed. Eng. 7, 50–72.

Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. NeuroImage 23, 806–826.

Schalk, G., Leuthardt, E., 2011. Brain-computer interfaces using electrocorticographic signals. IEEE Rev. Biomed. Eng. 4, 140–154.

Schmidt, D. M., George, J. S., Wood, C. C., 1999. Bayesian inference applied to the electromagnetic inverse problem. Hum. Brain Mapp. 7, 195–212.

Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., Wolpaw, J. R., 2006. A {P300} event-related potential brain–computer interface (bci): The effects of matrix size and inter stimulus interval on performance. Biol. Psychol. 73 (3), 242 – 252.

Sellers, E. W., Vaughan, T. M., Wolpaw, J. R., 2010. A brain-computer interface for long-term independent home use. Amyotroph. Lateral Scler. 11 (5), 449–455.

Serences, J. T., Yantis, S., 2006. Selective visual attention and perceptual coherence. Trends Cogn. Sci. 10, 38–45.

Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., Donoghue, J. P., 2002. Brain-machine interface: Instant neural control of a movement signal. Nature 416, 141–142.

Shen, J., Hu, H., 2006. Visual navigation of a museum guide robot. In: 6th World Congr. Intell. Control Autom. Vol. 2. pp. 9169–9173.

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M., 2011. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science 334 (6061), 1413–1415.

Shibata, K., Yamagishi, N., Goda, N., Yoshioka, T., Yamashita, O., Sato, M., Kawato, M., 2008. The effects of feature attention on prestimulus cortical activity in the human visual system. Cereb. Cortex. 18, 1664–1675.

Shimokawa, T., Kosaka, T., Yamashita, O., Hiroe, N., Amita, T., Inoue, Y., Sato, M., 2012. Hierarchical Bayesian estimation improves depth accuracy and spatial resolution of diffuse optical tomography. Opt. Express 20 (18), 20427–20446.

Shin, Y., Lee, S., Lee, J., Lee, H.-N., 2012. Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems. J. Neural Eng. 9 (5).

Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N. K., Ugurbil, K., 2007. Spatio-temporal point-spread function of fmri signal in human gray matter at 7 tesla. NeuroImage 35 (2), 539 – 552.

Sibley, G., Sukhatme, G. S., Matthies, L., 2007. Constant time sliding window filter slam as a basis for metric visual perception. In: IEEE Int. Conf. Robot. Autom. Workshop.

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., Engel, A. K., 2008. Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. Neuron 60, 709–719.

Simpson, R. C., LoPresti, E. F., Cooper, R. A., 2008. How many people would benefit from a smart wheelchair? J. Rehabil. Res. Dev. 45 (1), 53–71.

Sirotin, Y. B., Hillman, E. M. C., Bordier, C., Das, A., 2009. Spatiotemporal precision and hemodynamic mechanism of optical point spreads in alert primates. Proc. Natl. Acad. Sci. 106 (43), 18390–18395.

Sitaram, R., Caria, A., Birbaumer, N., 2009. Hemodynamic brain–computer interfaces for communication and rehabilitation. Neural Netw. 22 (9), 1320–1328.

Strangman, G., Culver, J. P., Thompson, J. H., Boas, D. A., 2002. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage 17, 719–731.

Sylvester, C. M., Shulman, G. L., Jack, A. I., Corbetta, M., 2007. Asymmetry of anticipatory activity in visual cortex predicts the locus of attention and perception. J. Neurosci. 27, 14424–14433.

Tan, D. S., Anton, N., 2010. Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction. Springer.

Tanaka, K., Matsunaga, K., Wang, H. O., 2005. Electroencephalogram-based control of an electric wheelchair. IEEE Trans. Robot. 21, 762–766.

Taylor, D. M., Tillery, S. I. H., Schwartz, A. B., 2002. Direct cortical control of 3d neuroprosthetic devices. Science 296 (5574), 1829–1832.

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D., 1999. Minerva: a second-generation museum tour-guide robot. In: IEEE Int. Conf. Robot. Autom. Vol. 3. pp. 1999–2005 vol.3.

Thut, G., Nietzel, A., Brandt, S. A., Pascual-Leone, A., 2006. α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J. Neurosci. 26, 9494–9502.

Tissari, S., Rahola, J., 2003. Error analysis of a galerkin method to solve the forward problem in MEG using the boundary element method. Comput. Methods Programs Biomed. 72 (3), 209 – 222.

Toda, A., Imamizu, H., Kawato, M., Sato, M., 2011. Reconstruction of two-dimensional movement trajectories from selected magnetoencephalography cortical currents by combined sparse Bayesian methods. NeuroImage 54, 892–905.

Tomasi, D., Ernst, T., Caparelli, E. C., Chang, L., 2006. Common deactivation patterns during working memory and visual attention tasks: An intra-subject fMRI study at 4 tesla. Human Brain Mapping 27 (8), 694–705.

Tongprasit, N., Kawewong, A., Hasegawa, O., 2011. Pirf-nav 2: Speeded-up online and incremental appearance-based slam in an indoor environment. In: IEEE Workshop Appl. Comput. Vis. pp. 145–152.

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., Dale, A. M., 1998. The retinotopy of visual spatial attention. Neuron 21, 1409–1422.

Uutela, K., Hamalainen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage 10, 173–180.

Van Gerven, M., Bahramisharif, A., Heskes, T., Jensen, O., 2009. Selecting features for BCI control based on a covert spatial attention paradigm. Neural Netw. 22, 1271–1277.

Vaughan, T., McFarland, D., Schalk, G., Sarnacki, W., Krusienski, D., Sellers, E., Wolpaw, J., 2006. The wadsworth bci research and development program: at home with bci. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2), 229–233.

Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., Schwartz, A. B., 2008. Cortical control of a prosthetic arm for self-feeding. Nature 453, 1098–1101.

Villringer, A., Chance, B., 1997. Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci. 20 (10), 435–442.

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, 101–104.

Virtanen, J., Noponen, T., Merilainen, P., 2009. Comparison of principal and independent component analysis in removing extracerebral interference from near-infrared spectroscopy signals. J. Biomed. Opt. 14, 054032.

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., Mehring, C., 2008. Hand movement direction decoded from MEG and EEG. J. Neurosci. 28, 1000–1008.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification. In: IEEE Conf. Comput. Vis. Pattern Recognit. pp. 3360–3367.

Wang, J. Z., Williamson, S. J., Kaufman, L., 1992. Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. IEEE Trans. Biomed. Eng. 39, 665–675.

Wästlund, E., Sponseller, K., Pettersson, O., 2010. What you see is where you go: Testing a gaze-driven power wheelchair for individuals with severe multiple disabilities. In: Symp. Eye-Tracking Res. Appl. (ETRA). pp. 133–136.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., Nicolelis, M. A. L., 2000. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 408, 361–365.

Wolf, M., Ferrari, M., Quaresima, V., 2007. Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. J. Biomed. Opt. 12 (6), 062104–062104–14.

Wolpaw, J., Wolpaw, E. W., 2012. Brain-Computer Interfaces: Principles and Practice. Oxford University Press.

Wolpaw, J. R., McFarland, D. J., 2004. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. 101 (51), 17849–17854.

Wolters, C. H., Anwander, A., Tricoche, X., Weinstein, D., Koch, M. A., MacLeod, R. S., 2006. Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: A simulation and visualization study using high-resolution finite element modeling. NeuroImage 30, 813–826.

Worden, M. S., Foxe, J. J., Wang, N., Simpson, G. V., 2000. Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-band electroencephalography increases over occipital cortex. J. Neurosci. 20, RC63.

Wu, J., Srinivasan, R., Kaur, A., Cramer, S. C., 2014. Resting-state cortical connectivity predicts motor skill acquisition. NeuroImage 91 (0), 84–90.

Wyart, V., Baudry, C. T., 2008. Neural dissociation between visual awareness and spatial attention. J. Neurosci. 28, 2667–2679.

Yamashita, O., Sato, M., Yoshioka, T., Tong, F., Kamitani, Y., 2008. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. NeuroImage 42, 1414–1429.

Yanagisawa, T., Hirata, M., Saitoh, Y., Kishima, H., Matsushita, K., Goto, T., Fukuma, R., Yokoi, H., Kamitani, Y., Yoshimine, T., 2012. Electrocorticographic control of a prosthetic arm in paralyzed patients. Annu. Neurol. 71, 353–361.

Yang, J., Yu, K., Gong, Y., Huang, T., 2009. Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conf. Comput. Vis. Pattern Recognit. pp. 1794–1801.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., Jang, J., 2009. NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. NeuroImage 44, 428–447.

Yoshimura, N., DaSalla, C. S., Hanakawa, T., Sato, M., Koike, Y., 2012. Reconstruction of flexor and extensor muscle activities from electroencephalography cortical currents. NeuroImage 59, 1324–1337.

Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., Sato, M., 2008. Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals. NeuroImage 42, 1397–1413.

Yuan, P., Gao, X., Allison, B., Wang, Y., Bin, G., Gao, S., 2013. A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces. J. Neural Eng. 10 (2), 026014.

Zhang, Q., Li, B., 2010. Discriminative K-SVD for dictionary learning in face recognition. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). pp. 2691–2698.

Zhou, M., Chen, H., Paisley, J., Ren, L., Sapiro, G., Carin, L., 2009. Non-parametric Bayesian dictionary learning for sparse image representations. In: Adv. Neural Inf. Process. Syst. Vol. 22.

Zhou, W., Yang, Y., Yu, Z., 2012. Discriminative dictionary learning for EEG signal classification in brain-computer interface. In: Int. Conf. Control Autom. Robot. Vis. (ICARCV) 2012. pp. 1582–1585.

List of Publications

Journals (Refereed)

Morioka, H., Kanemura, A., Morimoto, S., Yoshioka, T., Oba, S., Kawanabe, M., and Ishii, S. “Decoding spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior information." *NeuroImage*. Vol. 90, pp. 128–139, 2014. doi: 10.1016/j.neuroimage.2013.12.035

Morioka, H., Kanemura, A., Hirayama, J., Shikauchi, M., Ogawa, T., Ikeda, S., Kawanabe, M., Ishii, S. “Learning a common dictionary for subject-transfer decoding with resting calibration." *NeuroImage*. Vol. 111, pp. 167–178, 2015. doi: 10.1016/j.neuroimage.2015.02.015.

International Conferences (Refereed)

Kanemura, A., Morales, Y., Kawanabe, K., Morioka, H., Kallakuri, N., Ikeda, T., Miyashita, T., Hagita, N., and Ishii, S. “A waypoint-based framework in brain-controlled smart home environments: brain interfaces, domotics, and robotics integration." *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2013)*, pp. 865—870, Tokyo, Nov. 2013.

Morioka, H., Yi, S., and Hasegawa, O. “Vision-based mobile robot’s SLAM and navigation in crowded environments." *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2011)*, pp. 3998—4005, San Francisco, Sep. 2011.

Reviews

森岡博史，長谷川修，“人混みでも環境地図を学習して稼働する自律移動ロボットを開発," 画像ラボ，2011年7月号.

Non-Refereed Conferences/Workshops

Morioka, H., Kanemura, A., Hirayama, J., Shikauchi M., Ogawa, T., Ikeda, S., Kawanabe, M., and Ishii, S. “Subject-transfer decoding by learning a common dictionary from multisubject dataset." *10th AEARU Workshop on Computer Science and Web Technology*. Tsukuba, Feb. 2015.

Ogawa, T., Gupta, K. P., Yano, K., Abdur Rahim, J. A., Morioka, H., Hirayama, J., Yamaguchi, S., Ishikawa, A., Inoue, Y., Kawanabe, M., and Ishii, S. “Decoding daily behaviors from NIRS signatures by using a portable NIRS device in the daily-life environment." *Society for Neuroscience Annual Meeting (SfN2014)*. Washington, D.C., Nov. 2014.

Morioka, H., Kanemura, A., Hirayama, J., Ogawa, T., Ikeda, S., Kawanabe, M., and Ishii, S. “Subject-independent decoding from EEG through sparse learning of spatial bases common across sessions and subjects." *INCF Japan Node International Workshop: Advances in Neuroinformatics 2014 (AINI 2014)*. Wako, Sep. 2014.

Morioka, H., Kanemura, A., Hirayama, J., Shikauchi M., Ogawa, T., Ikeda, S., Kawanabe, M., and Ishii, S. “Subject-independent BMI through sparse learning of spatial bases common across sessions and subjects." *Neuro2014*. Yokohama, Sep. 2014.

Ogawa, T., Gupta, K. P., Abdur Rahim, J. A., Yano, K., Morioka, H., Hirayama, J., Kawanabe, M., and Ishii, S. “Decoding daily-life behavioral signatures in the real environment: portable NIRS signal using behavior labels." *Neuro2014*. Yokohama, Sep. 2014.

宮戸岳，大羽成征，森岡博史，石井信．“深層神経学習網による脳波デコーディングについて." 第16回情報論的学習理論ワークショップ（IBIS2013）, 東京，2013年11月.

Morioka, H., Kanemura, A., Morimoto, S., Yoshioka, T., Kawanabe, M., and Ishii, S. “Decoding of spatial attention from cortical currents estimated from EEG with NIRS prior." *Neuro2013*. Kyoto, Jun. 2013.

Morioka, H. “Improving decoding accuracy of spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior." *Systems Neurobiology Spring School (SNSS)*, Kyoto, Mar. 2013.

森岡博史，兼村厚範，川鍋一晃，吉岡琢，森本智志，石井信．“NIRSを事前情報としEEGから推定された皮質電流からの空間注意のデコーディング." 脳と心のメカニズム第13回

冬のワークショップ, 北海道，2013 年 1 月.

Morioka, H. “Decoding of spatial attention from near-infrared spectroscopy." *Systems Neurobiology Spring School (SNSS)*, Kyoto, Mar. 2012.

Morioka, H. “Vision-based mobile robot’s SLAM and navigation in crowded environments." *Joint Workshop on Dynamical Systems and Control*, Shanghai China, Nov. 2011.

森岡博史, 李想揆, Tongprasit Noppharit, 長谷川修. “人の多い混雑な環境下での SLAM による移動ロボットのナビゲーション." 第 28 回日本ロボット学会学術講演会，1Q3-6，名古屋，2010 年 9 月.

Patents

長谷川修，森岡博史，トンプラシット・ノッパリット，李想揆. “地図生成装置，地図生成方法，移動体の移動方法，及びロボット装置". 出願番号 2010-209657，公開番号 2012-64131

Awards

10th AEARU Workshop on Computer Science and Web Technology (CSWT-2015) Best Poster Award. Feb. 2015.

ATR 表彰優秀研究賞．ネットワーク型ブレイン・マシン・インタフェースの一般生活環境への適用可能性の実証．国際電気通信基礎技術研究所．2013 年 3 月.

IEEE Robotics and Automation Society Japan Chapter Young Award (IROS2011). Sep. 2011.

Press releases

「念じて機器動かす」，日経産業新聞，2014 年 3 月 17 日掲載.

「念じるだけ 開くドア」，読売新聞，2013 年 11 月 7 日掲載.

「念じて家電操作，ビックデータ活用課題」，日本経済新聞，2013 年 8 月 6 日掲載.

子供の科学-移動ロボットの紹介記事．2010 年 12 月号.

「景色で学ぶロボット！」，ワールドビジネスサテライト，2010 年 10 月 18 日.

「東工大，人混みの中でも SLAM が行える手法を展示，生活環境で稼働するロボの実現に」，Robonable，2010 年 10 月 1 日.

「東工大，人込みも最短で移動するロボ開発」，日刊工業新聞，2010 年 9 月 22 日.

「人混みでも環境地図を学習して稼働する自律移動ロボットを開発」，東京工業大学広報センター，2010 年 9 月 21 日.