2023

Decoding spatial location of attended audio-visual stimulus with EEG and fNIRS

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

DECODING SPATIAL LOCATION OF ATTENDED AUDIO-VISUAL STIMULUS WITH EEG AND FNIRS

by

MATTHEW H. NING

B.A., University of Rochester, 2015
M.S., Boston University, 2020

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2023

Approved by

First Reader

Kamal Sen, PhD
Associate Professor of Biomedical Engineering

Second Reader

Laura Lewis, PhD
Assistant Professor of Biomedical Engineering

Third Reader

Meryem A. Yücel, PhD
Research Associate Professor of Biomedical Engineering

Fourth Reader

Cara E. Stepp, PhD
Professor of Speech, Language and Hearing Sciences
Professor of Biomedical Engineering
Associate Professor of Otolaryngology

Fifth Reader

Ross K. Maddox, PhD
Associate Professor of Biomedical Engineering
Associate Professor of Neuroscience
University of Rochester

Acknowledgments

It all begins when I emailed Antje Ihlefeld, formerly at NJIT, for research opportunity. Since then, she took me in her lab and showed me the vast and highly interdisciplinary and wondrous world that is auditory neuroscience, and the people that makes up the community. I'm grateful to have witness, and being a part of, the endless cycle of reciprocal support and positive energy that flow between her and the community. Thank you to Antje and her former lab (Nima, Sahil and Pelin).

In the next chapter of my life at Boston University, which is nearing its end, I'm truly grateful to have the consistent support of my advisors Kamal Sen and Meryem Yucel. I'm grateful for Kamal's confidence in giving me this very important project. I'm thankful for Kamal's and Meryem's patience and their willingness to work with me in the beginning and the early part of the project. And I'm grateful that they stick with me to the end of the project. And I'm very glad that all of these is paying off at the end. When I saw a quote hanging in Meryem's office, "Be the change you want to see in the world", I know Meryem took it to her heart. And I have no doubt she'll continue to inspire the next generation of scientists from a diverse background. Lastly, thank you Kamal for calling me a brave man one time early in my PhD career.

In addition, I'm truly grateful for David Boas, Alex Von Luhmann, Ross Maddox, and Cara Stepp for their willingness to work with me, whom I have had the great privilege of learning from.

To those who lift me up, inspire me and challenge me to be greater, thank you to Kevin, Holly, John, Karina, Jiyoung, Chi, Sunit, Paul, Vasu, Kinyasi, Timothy, Jimmy, Rashaad, and Josephine. I'm truly grateful for your friendships and I'll never forgot all the crazy fun times we had together. I'm truly grateful for you showing me that the world is still a big place to explore and that there is so much fun, life and joy to be found. In addition, for a

few of you, I'm grateful for all of the conflicts we have gone through and resolved together. These only make us stronger.

Finally, I'm grateful for my parents and my sister who are the constant presences in my life and our family pet cat, Addie.

DECODING SPATIAL LOCATION OF ATTENDED AUDIO-VISUAL STIMULUS WITH EEG AND FNIRS

MATTHEW H. NING

Boston University, College of Engineering, 2023

Major Professors: Kamal Sen, PhD
Associate Professor of Biomedical Engineering

Meryem A. Yücel, PhD
Research Associate Professor of Biomedical Engineering

ABSTRACT

When analyzing complex scenes, humans often focus their attention on an object at a particular spatial location in the presence of background noises and irrelevant visual objects. The ability to decode the attended spatial location would facilitate brain computer interfaces (BCI) for complex scene analysis. Here, we tested two different neuroimaging technologies and investigated their capability to decode audio-visual spatial attention in the presence of competing stimuli from multiple locations. For functional near-infrared spectroscopy (fNIRS), we targeted dorsal frontoparietal network including frontal eye field (FEF) and intra-parietal sulcus (IPS) as well as superior temporal gyrus/planum temporal (STG/PT). They all were shown in previous functional magnetic resonance imaging (fMRI) studies to be activated by auditory, visual, or audio-visual spatial tasks. We found that fNIRS provides robust decoding of attended spatial locations for most participants and correlates with behavioral performance. Moreover, we found that FEF makes a large contribution to decoding performance. Surprisingly, the performance was significantly above chance level 1s after cue onset, which is well before the peak of the fNIRS response.

For electroencephalography (EEG), while there are several successful EEG-based algorithms, to date, all of them focused exclusively on auditory modality where eye-related artifacts are minimized or controlled. Successful integration into a more ecological typical usage requires careful consideration for eye-related artifacts which are inevitable. We showed that fast and reliable decoding can be done with or without ocular-removal algorithm. Our results show that EEG and fNIRS are promising platforms for compact, wearable technologies that could be applied to decode attended spatial location and reveal contributions of specific brain regions during complex scene analysis.

Contents

List of Tables

List of Figures

List of Abbreviations

AAD	.....	Auditory attendtion decoding algorithm
antIPS	.....	Anterior IPS
ASD	.....	Autism Spectrum Disorder
BCI	.....	Brain-Computer Interface
BOLD-fMRI	.....	Blood-oxygenation level dependent fMRI
CI	.....	Cochlear implant users
CSA	.....	Complex Scene Analysis
CSP	.....	Common Spatial Pattern
CW-fNIRS	.....	Continuous-Wave fNIRS
dFPN	.....	Dorsal Frontoparietal Network
dlPFC	.....	Dorsolateral Prefrontal Cortex
ECoG	.....	Electrocorticography
EEG	.....	Electroencephalography
EMG	.....	Electromyography
EOG	.....	Electrooculography
ERP	.....	Event-Related Potential
FEF	.....	Frontal Eye Field
fMRI	.....	Function Magnetic Resonance Imaging
fNIRS	.....	Functional Near-Infrared Spectroscopy
GEP	.....	Generalized Eigenvalue Problem
GLM	.....	General Linear Model
hEOG	.....	Horizontal EOG
HL	.....	Hearing Loss
HRF	.....	Hemodynamic Response Function
HRTF	.....	Head-Related Transfer Function
iEEG	.....	Intracranial Electroencephalography
iPCS	.....	Inferior Precentral Sulcus
IPS	.....	Intraparietal Sulcus
latIPS	.....	Lateral IPS
LOFO	.....	Leave-One-Feature-Out
MEG	.....	Magnetoencephalography
mEOG	.....	Extraocular muscles
MNI	.....	Montreal Neurological Institute

mSPL	. . . . .	Medial Superior Parietal Lobule
NH	. . . . .	Normal Hearing
OD	. . . . .	Optical Density
OLS	. . . . .	Ordinary Least Squares
PCA	. . . . .	Principal Component Analysis
PET	. . . . .	Positron-Emission Tomography
SNR	. . . . .	Signal-to-Noise Ratio
sPCS	. . . . .	Superior Precentral Sulcus
SPL1	. . . . .	Superior Parietal Lobule 1
SS	. . . . .	Short-Separation
STG/PT	. . . . .	Superior Temporal Gyurs/Planum Temporale
TA	. . . . .	Target-Alone
tgPSC	. . . . .	Transverse Gyrus intersecting Precentral Sulcus
TM	. . . . .	Target+Maskers
VHF	. . . . .	Very-High Frequency

Chapter 1

Introduction

1.1 Background

One important ability of the human brain that allows rich communication, interaction and more importantly, connections between humans is known as complex scene analysis (CSA). CSA is defined as selectively attending to specific objects in a complex scene with multiple objects. For example, at a shopping mall on a busy day, a normal-hearing individual is able to look at a friend and hear what they are saying in the midst of background noises and moving shoppers. Although this ability is usually taken for granted, there are still millions of people worldwide with varying degrees of hearing loss, autism spectrum disorder and ADHD that struggle with CSA. Thus, there is a need to design brain-computer interfaces (BCIs) and assistive devices that would address deficiency in CSA ability.

Complex scene analysis can be broken down into distinct processes: determining the spatial location of a target stimulus, segregating the target stimulus from the competing or background stimuli, and reconstructing the target stimulus from the mixture forming a perceptual object. There are already several algorithms that cover the segregation and reconstruction processes. For a review, see (Szabó, Denham, and Winkler 2016). However, all of them lack a critical information needed to complete the process: localization of spatial attention. We will use neuroimaging technologies with the goal of decoding the spatial attention. The five major types of functional and non-invasive neuroimaging technologies are EEG, MEG, fMRI, fNIRS and PET and will be briefly summarized here.

1.2 Comparison Between Neuroimaging Technologies

Electroencephalography (EEG) records electrical activities from the scalp with the main intent of recording electrical activities of the brain. Specifically, the main signal of interest in EEG recordings arises from extracellular currents which, in turn, arise mainly from synaptic currents. Electrical activities generated by an individual neuron are too small to be reliably detected by EEG. Thus EEG recordings likely reflect summation of synchronous activities from thousand to million of neurons with similar spatial orientation. Due to high conductivity, EEG recordings also capture signals not of interest and include electroocoulogram (EOG) and electromyogram (EMG) such as auricular and cardiac activities. EEG typically also captures electronics from the vicinity and power line noise from power source, resulting in degraded signal-to-noise ratio (SNR). It is also due to high conductivity that the spatial resolution is smeared and thus represents electrical activity at the macroscopic level. This effect is commonly known as volume conduction effect. Additionally, since the electric field gradient is inversely proportional to the square of distance, EEG signals from deep sources are sharply attenuated. The main advantages of EEG are the high temporal resolution, cost relative to other neuroimaging technologies, and portability. Its greater portability also allows EEG to be used in a wider range of experimental designs.

Magnetoencephalography (MEG) is intended to record the magnetic field of the brain. The main source of magnetic activity of neural origin is the net effect of ionic currents in dendrites that collectively act as current dipoles (Hämäläinen et al. 1993). MEG suffers from the same artifacts that EEG faces. However, since the magnetism generated by neuronal activities (usually measured in femtotesla) is smaller than the earth's magnetic field (usually measured in microtesla) roughly by the order of magnitude 9, two methods are employed to deal with this issue. Arrays of highly sensitive magnetometers, known as SQUID, short for superconducting quantum interference devices and immersed in liquid helium, are used. In addition, magnetic shielding is needed to reduce magnetic interference

from both the Earth and the surrounding. These two renders MEG as the most expensive out of 5 neuroimaging technologies listed. It is also due to SQUID and magnetic shielding that MEG is tied with fMRI and PET as the least portable neuroimaging technologies.

While EEG and MEG signals originate from the same underlying processes, there are still important distinctions to be made. First, due to difference in conductivity and permeability properties, magnetic fields are less distorted by the scalp and skull, thereby providing MEG slightly better spatial resolution. Second, MEG is more sensitive to tangential component of the detected current whereas EEG can detect both tangential and radial components of the detected current. Therefore, MEG usually detects neuronal activities from the sulci whereas EEG can detects neuronal activities from both sulci and gyri. Unlike electric field, magnetic dipole field diminish inversely proportional to the cube of the distance. Thus, MEG has better localization than EEG at the cost of more limited coverage area. In addition, MEG primarily detects intracellular currents whereas EEG primarily detect extracellular currents (S. Singh 2014).

CW-fNIRS, short for continuous-wave functional near-infrared spectroscopy, is an optical imaging technique that can be used to measure change in concentration of two different tissue chromophores: oxygenated and deoxygenated hemoglobin. Continuous-wave fNIRS emits light sources at fixed wavelengths and amplitudes and detect the same light scattered or reflected at detector sites. As its name implies, it adopts near-infrared lights (between 690 nm and 900 nm) due to their relatively high transmission in skin, bones and biological tissues. Among the few biological chromophores that absorb near-infrared light is hemoglobin, both oxygenated and deoxygenated. Since absorption spectra of oxygenated and deoxygenated hemoglobins differs, their relative concentration can be inferred. In the case of continuous wave fNIRS, modified Beer-Lambert law is used to approximately derive the change in concentration of hemoglobin (Delpy, 1988). Thus fNIRS measures the hemodynamic activities of the brain. The main drawback of CW-fNIRS is its differential

sensitivity to hair and skin pigmentation, making sampling subjects biased toward those with greater transmission to near-infrared light (less melanin). Its main advantages are lower cost and greater portability relative to the other hemodynamic-based neuroimaging technology, fMRI.

Functional magnetic resonance imaging (fMRI) is simply the application of nuclear magnetic resonance imaging to study the cortical activities. Very briefly, nuclear magnetic resonance operates by exciting and recording the energy (almost exclusively radio frequency in clinical and biomedical research setting) absorbed and emitted by a population of nuclei with nuclear spin property. Hydrogen is the conventional choice of targeted nuclei (although there exists carbon-based MRI). Thus, fMRI recordings represent the relative density of hydrogen nuclei, which are abundant in both water and fat. Blood oxygenation level dependent (BOLD)-fMRI, the most popular type of fMRI, is used to capture the hemodynamic response. During the hemodynamic response, there is a shift in the relative volume of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) in respond to an event or an activity. Since HbO and HbR have different magnetic susceptibility, BOLD-fMRI can exploit these differential paramagnetic properties. Just like every other neuroimaging technologies, fMRI recordings come with its own suite of artifacts. These include driftings induced by the instability of the applied magnetic field, spikes induced by motion-related artifacts and periodic signals induced by systemic physiological activities such as heart beating and respiration. Lastly, fMRI suffers from high cost and immobility relative to EEG and fNIRS.

fNIRS captures hemodynamic activities in all vascular compartments including arterioles, venules and capillaries (H. Liu et al. 1995) whereas BOLD-fMRI signals usually captures hemodynamic activities in both veins and capillaries (Drew 2019). In the case of both fNIRS and BOLD-fMRI, since they measure cerebral blood flow (CBF), their measurement of neural activities is indirect. The relationship between CBF and neural activities

is called neurovascular coupling. A canonical description of the neurovascular coupling is presented: when a region of the brain is engaged in a task, there is an associated increase in the demand for metabolic activities, including increase in CBF needed to supply neurons with oxygen and glucose (Heeger and Ress 2002). However, the oxygen consumption is much less than the oxygen supply delivered by CBF, resulting in net increase in oxygenation. This is being exploited by BOLD-fMRI, which detects change in deoxygenated hemoglobin. Later researches by (Logothetis 2002; Logothetis et al. 2001) demonstrated that BOLD increase reflects synaptic activities and intracortical processing rather than action potentials. Unlike fMRI, fNIRS has the advantage of being able to distinguish between oxy- and deoxygenated hemoglobin changes.

In addition, both fNIRS and fMRI suffer from periodic systemic physiological signals including cardiac pulsation and respiration which usually occur above 0.5 Hz and Mayer waves, which has a characteristic frequency of around 0.1 Hz.

PET, short for positron emission tomography, is an optical imaging technique. The patients are given radioactive substance (usually injected into the veins of the arms), called radiotracer, which eventually builds up in areas with high metabolic or biochemical activity and binds to test molecules. Upon binding, these radioactive substance emits gamma rays, which mainly arise from positrons. These gamma rays are eventually recorded by PET. PET is more commonly used in clinical settings, especially in the diagnosis of cancers and diseases. Ingestion of radioactive substance poses a small health risk. Thus it's considered to be minimally invasive. Like MEG and fMRI, it is very expensive and immobile. Its application in basic science remains restricted.

Lastly, for safety reason, cochlear implant (CI) users are not allowed to participate in fMRI and MEG experiments.

A comparison of the spatial and temporal resolution among 5 different neuroimaging technologies is shown in Figure 1·1

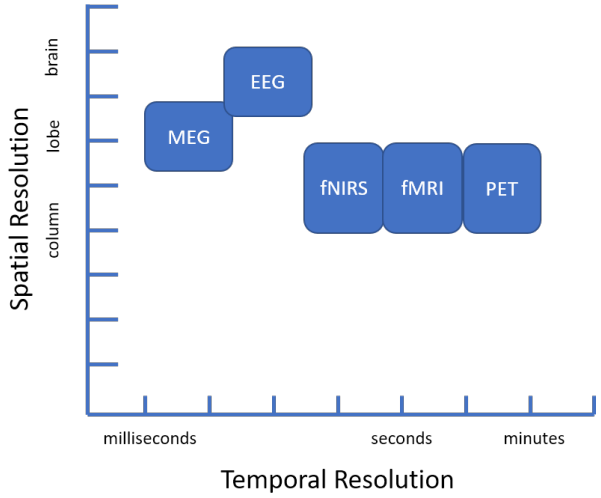

Figure 1·1: Spatial-temporal resolution of 5 different neuroimaging technologies. The axes are not drawn to scale.

Another factor to consider when comparing different neuroimaging technologies is the preparation time. fMRI and MEG have relatively fast preparation time whereas EEG and fNIRS have very slow preparation time because EEG and fNIRS require electrodes and optodes positioning, respectively. PET has the slowest preparation due to the time it takes for the radiotracer to reach to the targeted system and bind to test molecules. Finally, regarding penetration depth, fMRI and PET has the best penetration depth, followed by EEG. MEG and fNIRS are more sensitive to superficial layer of the cortex.

Since our priorities are cost and portability, we will test EEG and fNIRS separately for proof-of-concept.

1.3 EEG

There are several existing algorithms that can localize the spatial attention (Geirnaert, Vandecappelle, et al. 2021). Many of these studies employ electroencephalography (EEG). One important class of EEG-based auditory attention decoding (AAD) algorithms is the stimulus-reconstruction method (not to be confused with the reconstruction step of CSA), which uses the envelopes of EEG signals to reconstruct the attended speech envelope.

Briefly, the stimulus-reconstruction method calculated the correlation coefficients of the reconstructed speech envelope with the speech envelopes of individual speakers and select the one with the highest correlation coefficient as the attended speaker (Alickovic et al. 2019; Aroudi et al. 2019; Cheveigné et al. 2018; Mirkovic et al. 2015; Narayanan and A. Bertrand 2020; J. A. O'Sullivan et al. 2015; Taillez, Kollmeier, and Meyer 2020; Wong et al. 2018). Requiring access to the original speech envelopes presents an obvious problem to designing a real-time brain-computer interface (BCI) for AAD. A different class of AAD algorithms is based on EEG signals themselves and encompasses common spatial pattern (Geirnaert, Francart, and A. Bertrand 2021) and convolutional neural network (Ciccarelli et al. 2019; Su et al. 2022; Vandecappelle et al. 2021) and have demonstrated considerable successes. Specifically, they have shown them to be capable of fast and accurate decoding.

However, all of these AAD algorithms only employ auditory stimuli. Successful integration of EEG-based AAD into everyday usage requires careful consideration for ocular-related activities, which are inevitable. While there are several algorithms to deal with ocular-related activities, they can only reduce and not completely eliminate them. In addition, both visual and auditory cues play an important role in CSA. Psychophysical, neurophysiological and neuroimaging studies all have previously showed that visual and auditory modalities can influence each other. These studies suggest that a single auditory-visual object can be formed by drawing and merging features from visual and auditory stimuli (Bizley, Nodal, et al. 2007; Bizley, Maddox, and Lee 2016). Specifically, previous psychophysical studies showed that auditory features can be used to enhance visual object formation, which in turn, can enhance attention operating on visual object. This is known as cross-modal influence (Busse et al. 2005; Serences et al. 2005; Shomstein 2006). Cross-modal influence also entails visual features operating on auditory object and hence, attention on auditory object (Desimone and Duncan 1995; Knudsen 2007; Maddox et al. 2015). Cross-modal influences have been demonstrated via activation or modulation in primary visual

cortex by auditory stimuli (Petro, Paton, and Muckli 2017), single-unit and local field potential modulation in primary auditory cortex by visual stimuli (Bizley, Maddox, and Lee 2016; Kayser, Petkov, and Logothetis 2008) and functional magnetic resonance imaging (fMRI) studies on humans (Calvert, Brammer, et al. 1999; Calvert, Hansen, et al. 2001; Laurienti et al. 2002; Martuzzi et al. 2007). Thus, employing auditory-visual objects has the potential to enhance the brain's perception of the attended auditory-visual object and consequently, improve discriminability of spatial location of attended said objects.

In addition, the high temporal resolution of EEG enables us to study neural activities in the frequency domain. In particular, the six frequency bands (expressed in term of passband edge frequency) commonly recognized by EEG studies are delta (1.5-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), low-beta (13-17 Hz), high-beta (18-30 Hz) and gamma-low (30-40 Hz). In addition, we tested two additional frequency bands: gamma-medium (60-80 Hz) and gamma-high (100-150 Hz). Summary of each frequency band's roles will be discussed below:

Delta and theta. Delta and theta frequency-bands are linked to features of temporal structure of both auditory and visual stimuli during multimodal speech perception in various studies (Arnal and Giraud 2012; Lakatos et al. 2008; Park et al. 2015) For example, Gross et al., 2013 used MEG to show that the quasi-rhythmic components of speech entrains the phase of low-frequency (delta and theta) and the amplitude of high-frequency (gamma) oscillations in the auditory cortex (Gross et al. 2013). For a review, see (Biau and Kotz 2018).

Alpha. Previous EEG and MEG studies suggest alpha frequency-band to be involved in spatial attention in both auditory (Deng, Choi, and Shinn-Cunningham 2020; Frey et al. 2014; Klatt et al. 2018; N. Müller and Weisz 2012; Strauab, Wostmann, and Obleser 2014; Tune, Wöstmann, and Obleser 2018; Weisz et al. 2014) and visual modalities (Kelly, Lalor, et al. 2006; Thut 2006; Worden et al. 2000; Wöstmann et al. 2016). Anatomical

and physiological study on macaque and source reconstruction from MEG signals on human brain both show alpha and beta bands to be involved in feedback signaling pathway of both dorsal and ventral visual areas (Michalareas et al. 2016). This is further supported by alpha desynchronization observed in primary and secondary visual cortices of non-human primates (Buffalo et al. 2011; Kerkoerle et al. 2014; Xing et al. 2012), as well as studies on laminar projection patterns in non-human primate visual cortical areas (Barone et al. 2000; Goldman-Rakic and Rakic 1991; Markov et al. 2014). In addition, when subjects focus cued attention to a visual stimulus at the left or right spatial location, alpha power exhibits lateralized spatial pattern, with stronger suppression over the hemisphere contralateral to the attended location (Bengson, Mangun, and Mazaheri 2012; Capotosto et al. 2009; Grent-'t-Jong et al. 2011; Kelly, Gomez-Ramirez, and Foxe 2009; Y. Liu et al. 2014; Rajagovindan and Ding 2011; T. Rihs, C. Michel, and Thut 2009; T. A. Rihs, C. M. Michel, and Thut 2007; Sauseng, Klimesch, et al. 2005; Sauseng, Feldheim, et al. 2011; Thut 2006; Worden et al. 2000). Similar lateralization for both alpha and beta powers was also observed in auditory spatial top-down attention study (Thorpe, D'Zmura, and Srinivasan 2012). One study shows lateralization of alpha and beta powers of auditory spatial attention to be dependent on the relative spatial distance between the target and masker stimuli (Mock et al. 2015). However, several auditory spatial cueing studies show alpha lateralization over posterior regions, a spatial distribution pattern that is remarkably similar to those observed in visual attention studies (Ahveninen et al. 2013; Banerjee et al. 2011; Kerlin, Shahin, and Miller 2010; Thorpe, D'Zmura, and Srinivasan 2012). One possible explanation is that the absence of visual stimuli or the instruction to maintain fixed gaze do not necessarily prevent the eyes and the visual system from engaging or responding (Braga et al. 2016; Popov et al. 2022). Another proposed hypothesis is the existence of a higher-order and potentially supramodal central control system that function in regulating and delegating attention to sensory-specific system (auditory, visual, somatosensory, etc)

(Farah et al. 1989; Spagna, Mackie, and Fan 2015).

Beta. While alpha and beta are implicated in several shared roles including spatial attention, their spectral characteristics and modulation differ. In two ERP studies on visual attention by (Cómez et al. 1998; Vázquez Marrufo et al. 2001), power spectral density plots reveal a prominent peak in the alpha range and follow a sharp decay at frequency above alpha. Furthermore, interaction between attention and location factors induced a statistically significant decrease in alpha power and an increase in lower beta power during attention. Like alpha, beta also exhibited lateralization pattern in auditory spatial and top-down attention study by (Thorpe, D'Zmura, and Srinivasan 2012). Similar EEG and MEG studies were conducted for the beta frequency-band and showed them to be involved in top-down goal-driven but non-spatial attention in various tasks (Bressler and Richter 2015; Donner and Siegel 2011; Gao et al. 2017; Hanslmayr et al. 2007; Saarinen et al. 2015; Todorovic et al. 2015; Womelsdorf and Fries 2007). For a review on lower-beta, see (Biau and Kotz 2018). Finally, there are conflicting studies that inquire the necessity of further refinement of the beta frequency-band into high and low-beta. A big part of the variety in the specific ranges within beta lies in the differences in the stimuli presented, the tasks performed, the type of attention or perception studied, the measurement tools employed, the signal processing steps applied, the behavioral conditions being compared, using phase-locked vs non-phase-locked, etc. Another major factor is using the a priori definition of the beta range before the start of the experiment instead of looking at the entire spectral range. For studies specifically on visual spatial attention, (Vázquez Marrufo et al. 2001) defined the beta range a priori to be 15-17 Hz in their EEG study whereas (Cómez et al. 1998) measured the observed beta modulation to be around 15-17 Hz. Variability in the specific beta ranges extended to non-spatial attention studies. For example, in a "paired stimulus" sensory gating experimental paradigm, which was designed to evoke bottom-up stimulus-driven attention, the a priori beta range investigated is 12-20 Hz (Kisley and

Cornwell 2006). However, in a MEG study on auditory temporal prediction conducted by (Arnal, Doelling, and Poeppel 2015), they used time-frequency analysis to look at a broad range of frequency and the ranges of alpha-beta and beta bands observed are 10-14 Hz and 18-22 Hz, respectively. Similarly, two MEG studies by (Fujioka et al. 2012; Fujioka et al. 2009) investigated a broad range of frequency using time-frequency analysis and found modulation mainly in 20-22 Hz during presentation of auditory tone sequence. (Yamagishi et al. 2003) also used time-frequency analysis tool and reported modulation to be 13-21 Hz range in their MEG study. Studies with frequency above 22 Hz are rarely reported and even so, are usually reported together with the low beta. For example, EEG study that showed statistically significant differences between selective and non-selective attention during CSA looked at the entire pre-defined beta range (12-30 Hz) (Gao et al. 2017). However, an EEG study (Giannitrapani 1971) specifically found increase in high-beta activity (21-33 Hz) during presentation of unstructured stimuli (diffused visual pattern in occipital area and white noise in prefrontal area). Similarly, increase in high-beta activity (roughly 20-24 Hz) was observed in trials with faster reaction time compared with slower reaction time during presentation of either auditory or visual stimuli (Kamiński et al. 2012). With no perfect and fully scientific method to reconcile differences in reported frequency ranges of the beta frequency band, we simply divided the beta frequency-band into low-beta (13-17 Hz) and high-beta (18-30 Hz). Lastly, we hypothesized the low-beta frequency band to be the most intact from both ocular-related activities and EMG.

Gamma-Low. Both EEG (Gruber 1999) and MEG (Bauer, Stenner, et al. 2014; Bauer, Kluge, et al. 2012; Magazzini and K. D. Singh 2018; T. R. Marshall, O'Shea, et al. 2015; T. R. Marshall, Bergmann, and Jensen 2015; Siegel et al. 2008) studies show gamma frequency band to be involved in visual spatial attention, typically with increased power in the occipital lobe contralateral to the attended visual hemifield. For auditory spatial attention, while EEG (Senkowski et al. 2005) and MEG (Tiitinen et al. 1993) studies show that

gamma frequency band response is greater in attended versus unattended condition, later EEG study show weak evidence of lateralization of the gamma frequency band (Thorpe, D'Zmura, and Srinivasan 2012). However, multisensory studies on spatial attention suggest that audio-visual stimuli will elicit greater response in gamma frequency band than audio-alone or visual-alone stimuli (Sakowitz et al. 2001; Senkowski et al. 2005). Gamma is also shown to play a role in other types of attention to or perception of auditory or visual stimulus. For example, MEG study reported transient (100 ms) evoked modulation in neural activity at around 40 Hz in response to a train of auditory stimuli (Joliot, Ribary, and Llinás 1994). Similarly, combined EEG and MEG study using visual stimulus on humans showed phase-locked 40 Hz magnetic response in the occipital lobe and non-phase locked 40 Hz electrical response in several sites, with the maximum responses in the anterior (around Cz) and bilateral posterior (around T5 and T6) (Tallon-Baudry et al. 1997). Several more ERP studies corresponding to the gamma frequency band show conflicting results in the specific nature of modulation, with increase in gamma power in (Gurtubay et al. 2001; Haig, De Pascalis, and Gordon 1999; Watanabe et al. 2002) and decrease in gamma power in (O. Bertrand et al. 1998; Fell, Hinrichs, and Röschke 1997; L. Marshall, Mölle, and Bartsch 1996). Several authors suggest that these attention are specifically bottom-up stimulus-driven (Chalk et al. 2010; Fries 2015; Jia, Tanabe, and Kohn 2013; Michalareas et al. 2016; Richter et al. 2017). A full review on gamma activities can be found in (Kaiser and Lutzenberger 2003; Muthukumaraswamy 2013; Pulvermüller et al. 1997; Tallon-Baudry 1999). Regardless of the specific nature of modulation, gamma is an excellent candidate for our decoding algorithm.

Gamma-Medium and High. EEG/MEG literature on oscillation frequency beyond 40 Hz related to auditory or visual attention is rather sparse. For example, when subjects are tasked to attend to visual stimuli with specific features such as color, Muller and Keil observed broadband response in 55-70 Hz range (M. M. Müller and Keil 2004). When studies

with focus on attention are not available, studies on perception will be used instead. For example, one auditory perception and short-term memory-based ERP study using Go/No-Go task showed statistically significant task-effect in high frequency range ($>$ 90 Hz) but not in the 30-90 Hz range (Lenz et al. 2008). One example of human EEG study on visual perception that attempted to control for EMG presented a single moving bar to the left visual hemifield and compared it with a control condition that consisted of two moving bars moving in the opposite direction (Muller et al. 1996). They found stronger response in the hemisphere contralateral to the side of visual stimulation and the response is rather broadband (40-100 Hz). Instead, VHF are more often studied with intracranial EEG (iEEG) or electrocorticography (ECoG), both of which are invasive (Crone et al. 2001; Sinai et al. 2009; Towle et al. 2008). Unlike the low gamma frequency band, which tend to show modulation specifically at 40 HZ, modulation at medium and high gamma frequency bands tend to be broadband. For example, an ECoG study by (Ray et al. 2008) show that the auditory cortical response to auditory stimuli was enhanced under selective attention in the 80-150 Hz range. While these (iEEG or ECoG) studies show correlations between very high frequency oscillation and auditory-related tasks, these signals tend to be sharply attenuated by the scalp and corrupted by noise in scalp EEG recordings (Crone et al. 2001; Jerbi et al. 2009). To the best of our ability, we didn't find any EEG study on auditory or visual spatial attention that show modulation in $>$40 Hz range. For literature reviews on gamma frequency band, see (Kaiser and Lutzenberger 2005) or Chapter 37 of (Schomer and Niedermeyer 2011).

While the electric field generated by corneo-retinal dipole can propagate to EEG recordings, the propagation and distortion of ocular artifact (saccade and blinkings) on EEG is mainly symmetrical. Hagemann and Naumann reported that the differential and asymmetrical effects are largely confined to delta and theta frequency bands. Except for the frontopolar sites, differential EEG signals in alpha frequency band are negligible and even

when detected, are likely to be of neural origin. Differential EEG signals in beta frequency band and above are virtually absent. Other studies (Gasser, Sroka, and Möcks 1985; Whitton, Lue, and Moldofsky 1978) are in good agreement on the neural origin of alpha and beta differential signals during eye movements. Furthermore, saccade can generate cortical response in the visual cortex, known as visual lambda response (Yagi 1981). In addition, signals from saccade and blinking are transient in nature.

After eye movement, fixed gaze at off-center points induced mainly DC offset primarily in frontopolar sites as shown by (Ai et al. 2016). In contrast, the DC offset is practically invisible in electrodes located medially, including Fz and Cz channels. Since eye gaze at off-center points mainly induced DC offset, the asymmetrical effect is practically invisible in frequency domain.

Electromyographic (EMG) activities have prominent peak between 50 and 150 Hz but overall have spectral activities in 20-200 Hz range (Pulvermüller et al. 1997).

Chapter 2

Decoding Attended Spatial Location during Complex Scene Analysis with fNIRS.

2.1 Introduction.

Due to the slow temporal dynamic of hemodynamic activities, fNIRS signals have significantly lower temporal resolution than EEG. Thus investigating fNIRS signals from frequency domain is likely limited. However, due to limited light scattering distance, fNIRS has better spatial resolution than EEG and thus enables better localization of the brain regions.

fMRI has provided insights into brain regions involved in auditory-spatial short-term memory (Michalka, Kong, et al. 2015). Specifically, auditory short-term memory tasks elicited bilateral hemodynamic responses in the transverse gyrus intersecting precentral sulcus (Michalka, Kong, et al. 2015; Michalka, Rosen, et al. 2016). fMRI also showed that with demanding auditory spatial short-term memory tasks visuospatial regions, e.g., intraparietal sulcus 2-4 and superior parietal lobule 1, can also be activated. In addition, non-visuospatial regions near these regions, e.g., anterior intraparietal sulcus (antIPS), lateral intraparietal sulcus (latIPS), and medial superior parietal areas, were also activated by the same tasks (Michalka, Kong, et al. 2015; Michalka, Rosen, et al. 2016). In separate studies, fMRI also showed STG/PT to be activated in sound-localization tasks (Deouell et al. 2007; Zwaag et al. 2011). In addition, dorsal frontoparietal region, including frontal eye field in the frontal cortex and posterior parietal cortex and intraparietal sulcus, are shown

to be activated by both audio-spatial and visuospatial tasks in fMRI studies (Shomstein 2006; Smith et al. 2010; C.-T. Wu et al. 2007) and simultaneous magnetoencephalography (MEG)/EEG study (Larson and Lee 2013). Although fMRI has revealed a wealth of information regarding brain regions activated during a wide variety of tasks, one limitation of fMRI is that it not portable, making it difficult to apply in BCIs and assistive devices.

An alternative technology that is both portable and has good spatial resolution is fNIRS, making it well suited for applications in BCIs and assistive devices, as well as revealing specific brain regions activated during tasks. Previously, fNIRS has been applied to classify different sound categories (Hong and Santosa 2016), to identify spatial locations of noise stimuli (Tian et al. 2021), characterize hemodynamic responses to varying auditory stimuli (Luke et al. 2021; Pollonini et al. 2014; Steinmetzger et al. 2020) and investigate informational masking (Zhang, Alamatsaz, and Ihlefeld 2021; Zhang, Mary Ying, and Ihlefeld 2018). However, to the best of our knowledge at the time of the writing, to date fNIRS has not been applied to decode auditory-visual spatial attention during CSA. Here we investigate the capability of fNIRS in such a decoding task. We find that fNIRS provides a robust ($\geq 70\%$ cross-validated accuracy for 2-class classification) platform for decoding attended spatial location. Surprisingly, improvement in decoding performance is faster than anticipated given the time-scales of the underlying fNIRS signals. Finally, we also find that FEF is a critical brain region for decoding attended spatial location during CSA.

2.2 Materials and Methods.

This study employs continuous wave fNIRS (CW6, Techen System) using 690nm and 830nm wavelengths, with 50 Hz sampling rate. We used a size 56cm Landmark Cap (EasyCap, Herrsching, Germany) for all subjects. By fitting the cap on a head model marked with EEG 10-20 reference points, we positioned and secured rubber grommets with panel holes to the cap to hold both source and detector optodes. Multiple channels are supported with

frequency multiplexing. The fNIRS recording software is synched to the stimulus presentation with Neurospec MMB70-817 triggerbox (Neurospec AG, Switzerland).

Dataset. The videos used in the experiment are from AVSpeech, a publicly available dataset (`https://looking-to-listen.github.io/avspeech/index.html`). For target stimuli, each subject was shown 6 different speakers, each speaker having 5 different clips, all showing the speakers' faces clearly. Each clip was shown at all spatial locations at equal frequencies. For masker stimuli, each subject was shown 12 different speakers which are separated from the pool of speakers designated for target stimuli. These 12 speakers also have 5 different clips. For both pools of targets and maskers, all are English speakers. Each subject was presented with different pools of speakers and maskers. For each video clip, the audio clip was extracted and passed through three different head-related transfer functions (HRTF) to create auditory virtualization, one for each location (center, 45° to the left, and 45° to the right). The HRTF filters are based on measurements using the KEMAR head model (`https://sound.media.mit.edu/resources/KEMAR.html`). For each location, the HRTF-filtered audio clip replaced the original audio waveform in the video clip. Thus, there is a total of 90 trials, 30 for each location. Stimuli delivery software was written with Psychtoolbox-3 library.

Participants. 12 adults with normal hearing (age 19-48) were recruited for this study in accordance with the Institutional Review Board of Boston University. A COVID-19 protocol was developed and strictly adhered to. Participants were screened to exclude those with neurological disorders. Participants were briefed and consented before partaking in this study and were compensated for their time.

Experiment. A computer with three monitors was used for the experiment. The monitors were located at the three locations: center, 45° to the left, and 45° to the right (Fig. 2·1). They all are equidistant from the center where the subject is sitting upright. The auditory stimuli were delivered via earphone (ER-1 Etymotic Research Inc.) with eartips

(E-A-RLink 3A Insert Eartips). This study adopted event-related design. For each trial, a 2 seconds long audio-visual cue was delivered in the form of a white cross against a black background and a 2 kHz pure tone linearly ramped in the first 0.5 s. This audio-visual cue appeared randomly at one of the 3 spatial locations, indicating the location of the target speaker. The cue was followed by 3 videos, one for each location, one of which is the target speaker and the 2 remaining are the maskers. It's followed by two multiple-choice questions, each question containing 5 possible choices. These multiple-choice questions are always displayed at the center location. The two questions are related to face identification and words identification, presented in that order. In the face identification task, the subjects are presented five different faces and is tasked with correctly identifying the face of the target speaker shown in the video. In the word identification task, the subjects are presented five different transcripts and is tasked with correctly identifying the words spoken by the target speaker in the video. Upon the completion of two questions, a blank black screen of jittered duration with uniform distribution between 14 and 16 seconds appeared. This is to increase the statistical power of general linear model (GLM) fitting by preventing fNIRS signals from synching with the timing of the trial. Thereafter, an instruction to press the space bar to begin the next trial is displayed at the center location. While the audio-visual cue and the video clip last 2 and 3 seconds respectively, the subjects have 20 seconds in total to answer both multiple-choice questions but can move on to the space bar instruction immediately after the questions are done. Lastly, the subjects are provided with chin rest in order to discourage head movements.

fNIRS probe design. The probe was designed in AtlasViewer (`https://github.com/BUNPC/AtlasViewer`), an open-sourced MATLAB software and is shown in Figure 2·2. The probe contains 12 sources and 17 regular-separation detectors and 6 short-separation detectors, for a total of 30 regular channels (30 mm) and 6 short-separation (SS) channels (8 mm). The probe covers the following regions: intraparietal sulcus (IPS) 2, 3

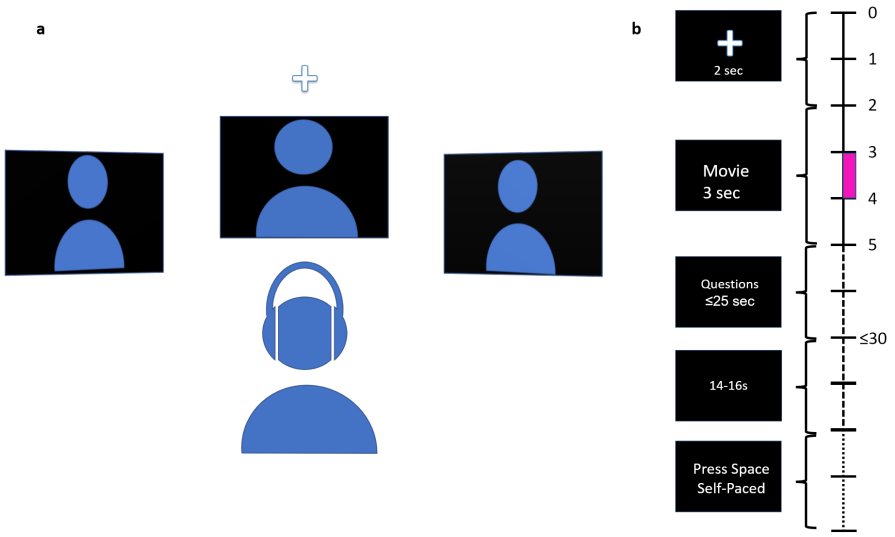

Figure 2·1: Schematic diagram of the experiment and trial. (a) Schematic diagram of experimental setup. Subject is seated in front of 3 spatially separated monitors, one at -45°to the left, one at 45°to the right, and one at 0°at the center. Audio is delivered via earphone. (b) Timeline of trial. The time at the right indicates the length of different parts of trial. A cue in the form of a white cross accompanied by a pure ramping tone of 2 kHz randomly appears at one of the three locations to indicate the location of the target stimuli at the start of the trial. Next, 3 video clips are displayed simultaneously for 3 seconds. Next, two multiple-choice questions are displayed at the center monitor, to be answered with keyboard. 1st question is to identify the face of the speaker, 2nd question is to identify the words spoken by the target speaker. Subject has up to 25 seconds to answer both questions. Upon the completion of the questions, it's immediately followed by a blank screen of jittered duration between 14-16 seconds. Next, instruction to press the space bar is displayed at the center monitor to begin the next trial. Pink box in $3 \leq t \leq 4$ is the decision window used in Figure 2·5, 2·6, 2·8

and 4, superior precentral sulcus (sPCS), inferior precentral sulcus (iPCS), dorsolateral prefrontal cortex (dlPFC), frontal eye field (FEF), posterior superior temporal gyrus/planum temporale (STG/PT), anterior intraparietal sulcus (antIPS), superior parietal lobule (SPL1) including medial superior parietal lobule (mSPL), and transverse gyrus intersecting precentral sulcus (tgPSC). The probe was designed with multiple channels to cover variation in reported MNI coordinates of FEF.

SNR. We excluded subjects that had at least 20 channels pruned using a cutoff of SNR = 1.5, where SNR is estimated as the mean divided by the standard deviation of the fNIRS signal.

Grouping by Task Performance. We divided subjects into 2 groups, the high-performing group consisted of subjects that scored at least 80% correct on trials, and the low-performing group consisted of subjects that scored less than 80%. A trial is counted as correct if the participant correctly identified both the faces of speakers and the words spoken by the targeted speakers.

fNIRS processing. For Fig. 2·3-2·9, raw light intensities were converted to optical densities using the mean of the signal as the arbitrary reference point. Next, motion artifacts in optical density (OD) were identified and corrected with targeted PCA before applying criteria for automatic rejection of stimulus trials (Yücel et al. 2014). Next, OD signals were band-pass filtered between 0.01 Hz and 0.5 Hz with a 3rd order zero-phase Butterworth filter. Next, the filtered OD were converted to chromophore concentration changes. The differential path length factor is held fixed at 1 and the concentration unit is Molar*mm (Scholkmann and Wolf 2013). Finally, systemic physiological signal clutter was regressed out using a GLM with short-separation channels modeling the signal clutter. Each regular channel was assigned a SS channel with the highest correlation Gagnon et al. 2011. All were done using the Homer3 open source fNIRS analysis package and custom MATLAB scripts.

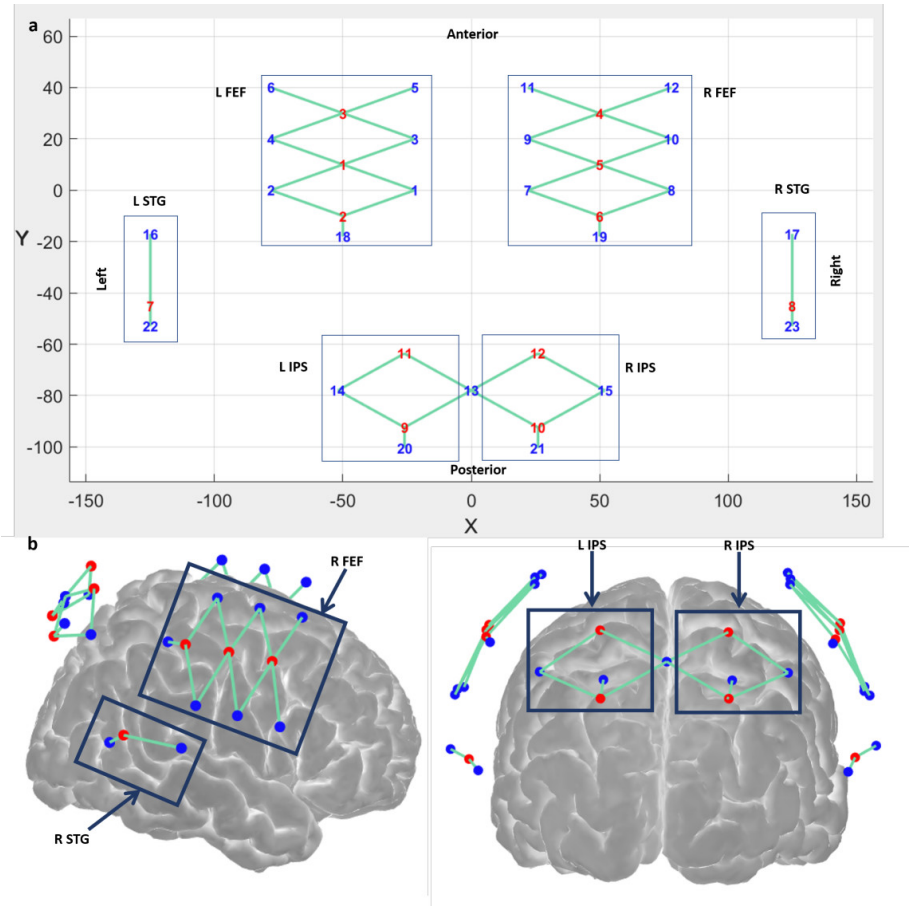

Figure 2·2: Probe Design. (a) 2D layout of probe design. The probe is subdivided into 5 regions: left frontal eye field (L FEF), right frontal eye field (R FEF), left posterior superior temporal gyrus (L STG), right posterior superior temporal gyrus (R STG), left intraparietal sulcus (L IPS) and right intraparietal sulcus (R IPS). Top side of the probe corresponds to the anterior side of the head, bottom is posterior. Red numbers represent source optoces, blue numbers represent detector optodes (b) 3D layout of probe design. Red circles represent source optodes, blue circles represent detector optodes, and turquoise lines represent channels.

For Fig. (Fig. 2·10), in order to accurately assess the latency of the classification accuracy, the OD signals were band-pass filtered between 0.01 Hz and 10 Hz with a 3rd order zero-phase Butterworth filter.

GLM. For hemodynamic response function (HRF) modeling, we fitted the GLM model to each subject. The GLM fitting here uses two classes of regressors: HRF for each condition and the systemic signal clutter. The temporal basis functions used to model the HRF is a sequence of 16 Gaussian functions ($b_1^{Conditioni}, \ldots, b_H^{Conditioni}, H = 16$), spaced 1 second apart (Δt=1s) and each with a width of 1 sec as we typically use (Gagnon et al. 2011) (σ = 1s). This flexible model offers better fitting of the HRF shape at the expense of more parameter estimations than the typical canonical hemodynamic response function (Lindquist et al. 2009). Short-separation (SS) fNIRS signals are used as physiological regressors. The usual assumptions of the GLM are held. For each short-separation channel y_{SS_i}, the GLM formulation in matrix format is as followed:

$$Y = A\beta + E \tag{2.1}$$

where $Y \in \mathbb{R}^{T \times N_{SS_i}}$ represents fNIRS recordings from all T time points and N_{SS_i} channels using y_{SS_i} channel, $\beta \in \mathbb{R}^{M \times N_{SS_i}}$ is the coefficient matrix with M regressors and N_{SS_i} channels, $E \in \mathbb{R}^{T \times N_{SS_i}}$ represents the residual/noise term and $A \in \mathbb{R}^{T \times M}$ is the design matrix and is defined as followed:

$$A = \begin{bmatrix} b^{Condition\ 1} & b^{Condition\ 2} & y_{SS_i} \end{bmatrix} \tag{2.2}$$

where $b^{Condition\ 1} \in \mathbb{R}^{T \times H}, Condition\ i \in \{1,2\}$ is a submatrix composed of the following column time vectors:

$$b^{Condition\ i} = \begin{bmatrix} b_1^{Condition\ i} & \ldots & b_H^{Condition\ i} \end{bmatrix} \tag{2.3}$$

and each column vector $b_1^{Condition\ i} \in \mathbb{R}^{T\times 1}, h \in \{1,\ldots,H\}$ is defined as followed:

$$b^{Condition\ i} = \sum_{k=1}^{K} b(t - \delta_k^{Condition\ i} - \Delta t \cdot h, \sigma) \tag{2.4}$$

where K is the total number of trials in Condition i and $\delta_k^{Condition\ i}$ represents the timing of the onset of k^{th} stimulus. The Gaussian function $b(\mu,\sigma)$ is defined as followed:

$$b(\mu,\sigma) = exp\big(\frac{-(x-\mu)^2}{2\sigma^2}\big), \tag{2.5}$$

where both μ and σ are in seconds. Letting $A^\top$ be the transpose of A. The β coefficients in the GLM equation are solved using ordinary least squares (OLS) method Ye et al. 2009:

$$\hat{\beta} = (A^\top A)^{-1} A^\top Y \tag{2.6}$$

For the group-level HRFs, we average HRFs of all subjects for each channel and spatial location.

For the classification, fitting the GLM regression model to the entire data first before cross-validating the trials would result in information leakage (Lühmann et al. 2020). In order to avoid leakage, we cross-validated both the GLM regression and classification steps. In each fold of the cross-validation, we fit the GLM model to the training dataset and estimated the coefficients using the OLS method. Then the SS coefficients $\hat{\beta}_{SS}$ estimated from the training set are used for the test set where the individual trials are the difference between the measured fNIRS signals and the systemic physiological regressor weighted by the SS coefficients.

$$\hat{y}_n^{test}(t) = y_n^{test}(t) - \hat{\beta}_{SS} \cdot y_{SS_i}^{test}(t) \tag{2.7}$$

Statistical Tests for GLM. For subject-level statistical test, to determine whether the hemodynamic activation significantly differed from 0, a one-tailed t-test was used on the estimated β-weights for the Gaussian temporal functions. We restricted ourselves to the

β-weights with peaks between 1 and 8 seconds from the stimulus onset. The t-statistic was computed using the ratio of the sum of β-weight estimators to the square root of the variance of the sum of β-weights estimators. To determine whether the activations between two different conditions (e.g., left vs right) differed, the t-statistic was computed using the following contrast vector:

$$t_{\hat{\beta}} = \frac{\mathbf{c}^\top \hat{\beta}}{\sqrt{Var(\varepsilon_n)\mathbf{c}^\top (A^\dagger A^{\dagger\top})\mathbf{c}}} \tag{2.8}$$

where ε_n is the error term for the n^{th} channel, $\mathbf{c} \in \mathbb{R}^{M\times 1}$ is a contrast vector for two different conditions, $A^\dagger \in \mathbb{R}^{M\times T}$ is the Moore-Penrose inverse of the design matrix, $\hat{\beta}_n \in \mathbb{R}^{M\times 1}$ are the estimated β weights for channel n. The Moore-Penrose inverse of the design matrix A assuming full column rank is:

$$A^\dagger = \left(A^\top A\right)^{-1} A^\top \tag{2.9}$$

Let $Var(\hat{\beta}_{Condition\ i,n}) \in \mathbb{R}^{H\times H}$ be the submatrix of $Var(\varepsilon_n)A^\dagger A^{\dagger\top}$ belonging to *Condition i* and channel n. Then $Var(\hat{\beta}_{Condition\ i,n})$ represents the covariance matrix for the HRF regressors for *Condition i* and channel n.

The 95% confidence intervals of the HRF for each condition and channel were computed as:

$$\hat{HRF}_{Condition\ i,n} \pm t_{97.5\%,T-(M+1)} \sqrt{\mathbf{diag}\left(\varphi Var(\hat{\beta}_{Condition\ i,n})\varphi^\top\right)}, \tag{2.10}$$

where we explicitly modeled independence between regressors by using only the diagonal elements of the matrix. $\varphi \in \mathbb{R}^{T_{HRF}\times H}$ is the matrix of the standard Gaussian temporal basis functions with T_{HRF} as the time vector for the HRF. Then φ is constructed as followed:

$$\varphi = \begin{bmatrix} b(\Delta t \cdot 1, 1) & \dots & b(\Delta t \cdot H, 1) \end{bmatrix} \tag{2.11}$$

$t_{\alpha,\nu}$ is the t-value of a student's t-distribution parameterized by the critical value α and the degree of freedom ν.

For group-level statistical test, one-sample two-tailed t-test is used to determine whether the activation differed from 0 and two-sample two-tailed t-test is used to determine whether the activations between two conditions differed. Instead of using β-weights estimators, we defined the average of the HRF values over the range [1,8] seconds as a single observation for each subject, channel and condition. Then a t-test is performed over a sample of averaged HRF values for each channel. Independence between β-weight estimators is assumed.

To account for the large variation in the shapes of the HRFs, we provide an alternative group-level statistical test. We call this the aggregation method and will be described as followed. For each channel and condition, we count the percentage of subjects that are statistically significant at significance level $\alpha = 0.05$, where equation (8) is used for statistical tests.

fNIRS classification. In this study, to test the cross-validation (CV) accuracy of the classifier trained with the entire probe, we tested 2-class classification between left (-45°) and right (45°) spatial locations as well as 3-class classification between all 3 spatial locations. This classification will be termed all-channel classification. The features used are the area under curve of one-second-long segment of fNIRS signals. The classifier used is linear discriminant analysis with linear shrinkage of covariance matrix (Ledoit and Wolf 2004). 5 other classifiers were also tested and their CV accuracies are reported in (Fig. 2·7). These are linear discriminant with non-linear estimator of covariance matrix using nuclear norm penalty (Chi and Lange 2014), K-Nearest Neighbors (Fix and Hodges 1989) with the distance metric defined as 1 minus cosine similarity and 10 nearest neighbors, support vector machine (Boser, Guyon, and V. N. Vapnik 1992; Cortes and V. Vapnik 1995; V. N. Vapnik 1997) using cubic polynomial kernel, regularized with ridge penalty, random

forest ensemble using 200 unpruned decision trees as weak learners, sampling $\sqrt{N}$ out of N channels for predictors and nonparametric bootstrapping (Tin Kam Ho 1995), and gentle adaptive boosting using 200 unpruned decision trees as weak learners (Friedman, Hastie, and Tibshirani 2000). First, we tested [ΔHbO], [ΔHbR] and [ΔHbT] on all subjects and reported the grand average of CV accuracies. Next, we focused on [ΔHbT] and divided the subjects into two groups based on their task performance and reported their CV accuracies. The CV accuracies are reported for the interval 3-4 s after the cue onset in (Fig. 2·5). We also plotted the correlations between subjects' task performance and individual subjects' CV accuracies using [ΔHbT] over interval 3-4 s after the cue onset (Fig. 2·6).

To assess the tradeoff between the number of channels, latency, and cross-validation accuracies, we plotted the CV accuracies of classifiers trained with different regions of the probe as a function of window length in Fig 2·10. All intervals tested started at cue onset. To account for the variation in reported MNI locations of different ROIs, as well as the variation in head and brain sizes and ROIs' actual locations, neighboring channels are aggregated to form 5 different subsets. For example, a subset named left FEF (L FEF) has 10 channels and covers FEF and the following nearby ROIs: iPCS, dlPFC, tgPCS, and sPCS in the left hemisphere. Similarly, right FEF (R FEF) has 10 channels and covers FEF and nearby ROIs in the right hemisphere, left STG subset has 1 channel and covers left STG region, right STG subset has one channel and covers right STG region, left IPS subset has 4 channels and covers IPS2-4, mSPL, antIPS and SPL1 in the left hemisphere and similarly for right IPS (Fig. 2·2).

To quantify the contribution of each channel, we excluded it from the all-channel classification and then used it to train a classifier and retrieve its CV accuracy. This will be called leave-one-feature-out (LOFO) classification. Specifically, we took the difference in CV accuracies between all-channel classification and LOFO classification and defined it as the LOFO impact (Dingcheng Feng, Feng Chen, and Wenli Xu 2013). Finally, we ran

single-channel classification where a single channel is used to train a classifier (Fig. 2·8). This is done for all channels. We used 10 repetitions of 5-fold cross validation in all cases except for linear discriminant with nonlinear estimator of covariance matrix, which used 10 repetitions of 5-fold nested cross validation. In all cases, 1-0 loss function was used. For the confidence interval of the means of cross-validation accuracy, we used a confidence level of 95% using student's t-distribution.

Statistical Tests for Classification. To test for statistical significance between 3 different chromophores, we used one-way ANOVA test. If the overall F test passed the given significance of 0.05, we followed up with post-hoc two-tailed two-sampled t-tests and used Bonferroni-corrected p-values. To test for statistical significance in decoding performance between high and low-performance group, we used two-tailed two-sampled t-tests.

2.3 Results

Behavioral Results. All 12 subjects performed significantly above chance level in the task (mean ± standard deviation: 84.3±14.1%). Of these 7 subjects, who had performance levels of 80% or above were categorized as high-performing, and 5 subjects with performance levels less than 80% were categorized as low performing.

HRFs. We plot the HRFs for various conditions, spatial locations and chromophores in the figures below. One example channel that shows different HRF patterns for left and right spatial locations is located in the FEF region in the right hemisphere and is shown in the right column of Figure 2·3. In addition, this channel has the highest correlation coefficient between task performance and single-channel CV accuracies. Its contralateral channel (in the left hemisphere) is also shown in the left column of Figure 2·3. In addition, we observed mostly a positive hemodynamic response in the right hemisphere and mostly a negative hemodynamic response in the left hemisphere in Δ[HbO] and Δ[HbT]. The different magnitudes of the responses to spatial location suggests that it's possible to decode the

spatial location using features of the HRF. In addition, we tested for statistical significance and only found that the Δ[HbO] activation in the right FEF in response to the left spatial location was statistically significant at $\alpha = 0.05$ $(t(11) = 2.73, p = 0.0196)$.

To investigate if there were any differences between the high-performance and low-performance groups, we examined the group-averaged HRFs. Figure 2·4 shows example group averaged HRFs from the FEF region. The most striking pattern was that the response variabilities (as shown with 95% confidence intervals) in the HRFs are substantially wider in the low-performance group compared to the high-performance group. Moreover, the positive hemodynamic response in the right hemisphere and negative hemodynamic response in the left hemisphere was more pronounced in the high-performance group. Statistical tests showed that only Δ[HbT] activation on the left FEF in response to the left spatial location in the high-performance group is statistically significant at $\alpha = 0.05, (t(6) = -2.64, p = 0.039)$.

Due to large subject variability and the small sample size, Table 2.1 provides the aggregation method where we calculated the percentage of subjects that achieved subject-level statistical significance using $\alpha = 0.05$. In all but one of the cases, subjects from the high-performance group are more likely to achieve statistical significance than subjects from the low-performing group.

High-Performance Group (n=7)							
ΔHbT	Left	Right	Center	**ΔHbT**	L vs R	L vs C	R vs C
L FEF	85.71	85.71	71.43	**L FEF**	85.71	71.43	57.14
R FEF	100.00	71.43	85.71	**R FEF**	71.43	85.71	100.00
Low-Performance Group (n=5)							
ΔHbT	Left	Right	Center	**ΔHbT**	L vs R	L vs C	R vs C
L FEF	60.00	40.00	40.00	**L FEF**	40.00	60.00	40.00
R FEF	80.00	40.00	60.00	**R FEF**	80.00	20.00	60.00

Table 2.1: Aggregation Method. Percentage of subjects with statistical significance for both activation (using one-sample one-tailed t-test) and difference in activation between two spatial locations (using two-sample one-tailed t-test). High-performance group is more likely to yield statistical significance in all except one case. All use significance of 0.05.

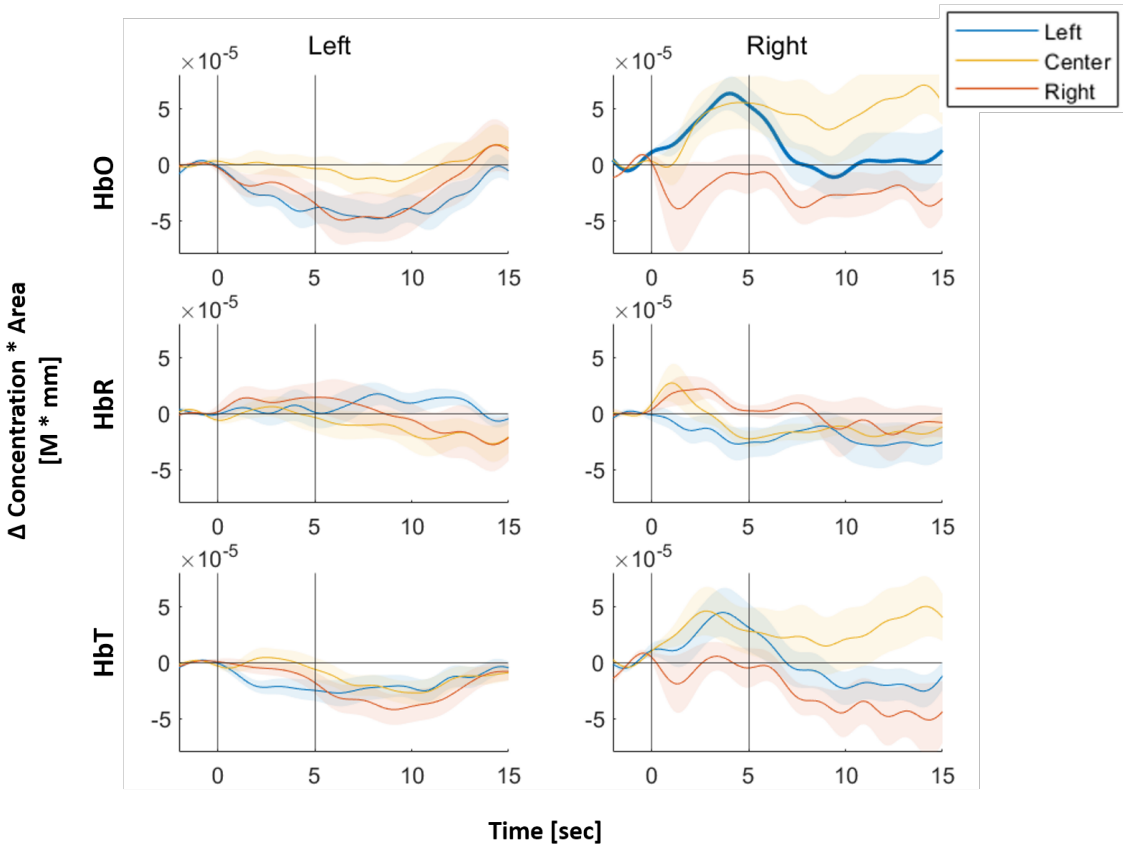

Figure 2·3: Hemodynamic Response Functions (HRFs) of FEF. Top panel represents Δ[HbO], middle panel represents Δ[HbR] and bottom panel represents Δ[HbT]. Left column represents channel covering FEF in the left hemisphere. Similar for right column. Blue line represents GLM-fitted HRF responses to left stimuli, red line for right stimuli, and yellow line for center stimuli. Shaded regions represent 95% confidence interval. Bold line indicates statistical significance in activation using one-sample two-tailed t-test. $*$ indicates statistical significance in difference between left and center condition using two-sample two-tailed t-test, + for left vs right condition, and x for right vs center condition. In this case, there is no statistical difference between any pairs of spatial locations. The horizontal black line is the axis at 0 $M \cdot mm$. Right FEF channel showed the highest correlation coefficient (R = 0.38, n = 9) between task performance and single-channel CV accuracies

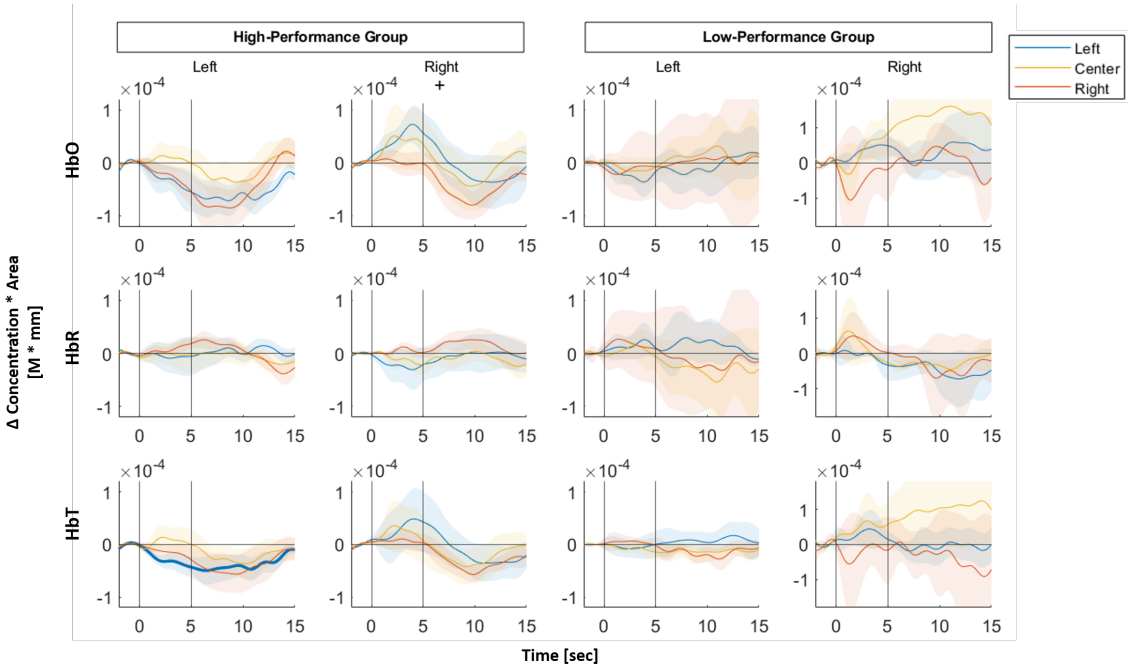

Figure 2·4: HRFs of FEF by task performance. Same as Figure 3 but for group average of a high and a low-performing groups. Range in M*mm of y-axis is [-1.2e-4 1.2e-4] for the 3 left columns but [-1.8e-4 1.8e-4] for the rightmost column.

Classification/Decoding Performance. Next, we plotted the decoding performance using all-channel cross validation (CV) accuracies. Figure 2·5a and c show the averages of 2-class and 3-class CV accuracies, respectively, for the interval t=3-4s after the cue onset when using all channels in the classification. In both cases, CV accuracies are significantly above chance level and are highest when using Δ[HbT] and lowest when using Δ[HbR]. Next, we focused on Δ[HbT] and divided the subjects into high-performance and low-performance groups. We found that decoding performance was significantly higher in high performance subjects compared to low performance subjects (Fig. 2·5b and d). To further quantify the relationship between CV accuracies and task performance, we plotted these quantities in Fig. 2·6a-b and performed linear regression when classifying using Δ[HbT] between 3 and 4 seconds. The significant correlation between the CV accuracies and task performance indicates that the decoding performance is dependent on the subject's task performance.

We next focused on CV accuracy in the high-performance group using Δ[HbT] signals. To assess the contribution of specific brain regions towards decoding performance we quantified single channel classification performance. Figure 2·8 shows the single-channel 2-class classification using a 2D channel layout (a) and interpolated over the brain surface (b) for the interval 3-4s after the cue onset. On average, decoding performance was higher in the right hemisphere but the difference was not significant at the 0.05 level (Fig. 2·8a inset). In addition, the top 2 performing channels are located in the right FEF region. The CV accuracy from single-channel classification (Fig. 2·8) was lower compared to all-channel classification (Fig. 2·5b).

In order to examine the contribution of each subset of channels to the CV accuracy as well as the tradeoff between the number of channels and window length, we plotted 2-class CV accuracies for left and right FEF and IPS regions (Fig. 2·10). All-channel classification is shown for comparison. All intervals start at the cue onset. For the all-channel classifica-

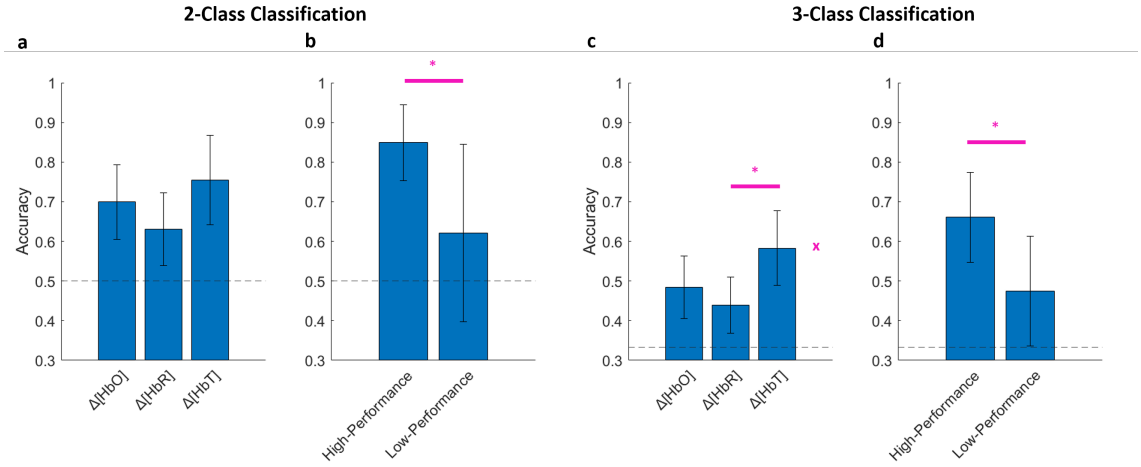

Figure 2·5: Δ[HbT] has the highest CV accuracies, CV accuracies are highly dependent on task performance. (a) 2-class grand-average CV results for all 12 subjects at t = 3-4s. Error bar represents 95% confidence interval. x represents F-test using one-way ANOVA with significance of 0.05. x represents significance level of $p < 0.05$. ($F_{2,33} = 1.86, p = 0.17$). (b) Focusing on Δ[HbT], 2-class CV accuracies shown for high-performance and low-performance groups. Two-sample two-tailed t-test: $t(10) = 2.79, p = 0.019$. Similar plots are shown for 3-class classification in (c) and (d). (c) One-way ANOVA F-test, $F_{2,33} = 3.87, p = 0.031$. (d) Two-sample two-tailed t-test: $t(10) = 2.69, p = 0.023$.

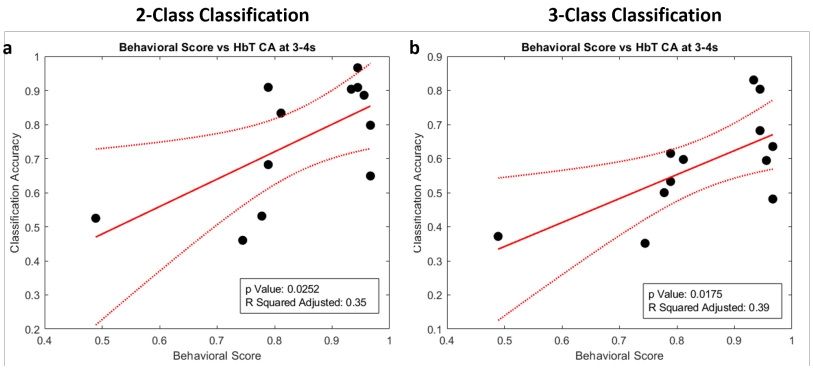

Figure 2·6: Strong correlation between CV accuracies and task performance. (a) Linear regression between behavioral score and 2-class CV accuracy for Δ[HbT] using interval 3-4s. (Slope coefficient: t(10) = 2.63, p=0.025, adjusted r2=0.35) (b) Linear regression between behavioral score and 3-class CV accuracy for Δ[HbT] using interval 3-4s. (Slope coefficient: t(10) = 4.84, p=0.018, adjusted r2=0.39). In all cases, n = 12. Dashed lines represent 95% confidence intervals.

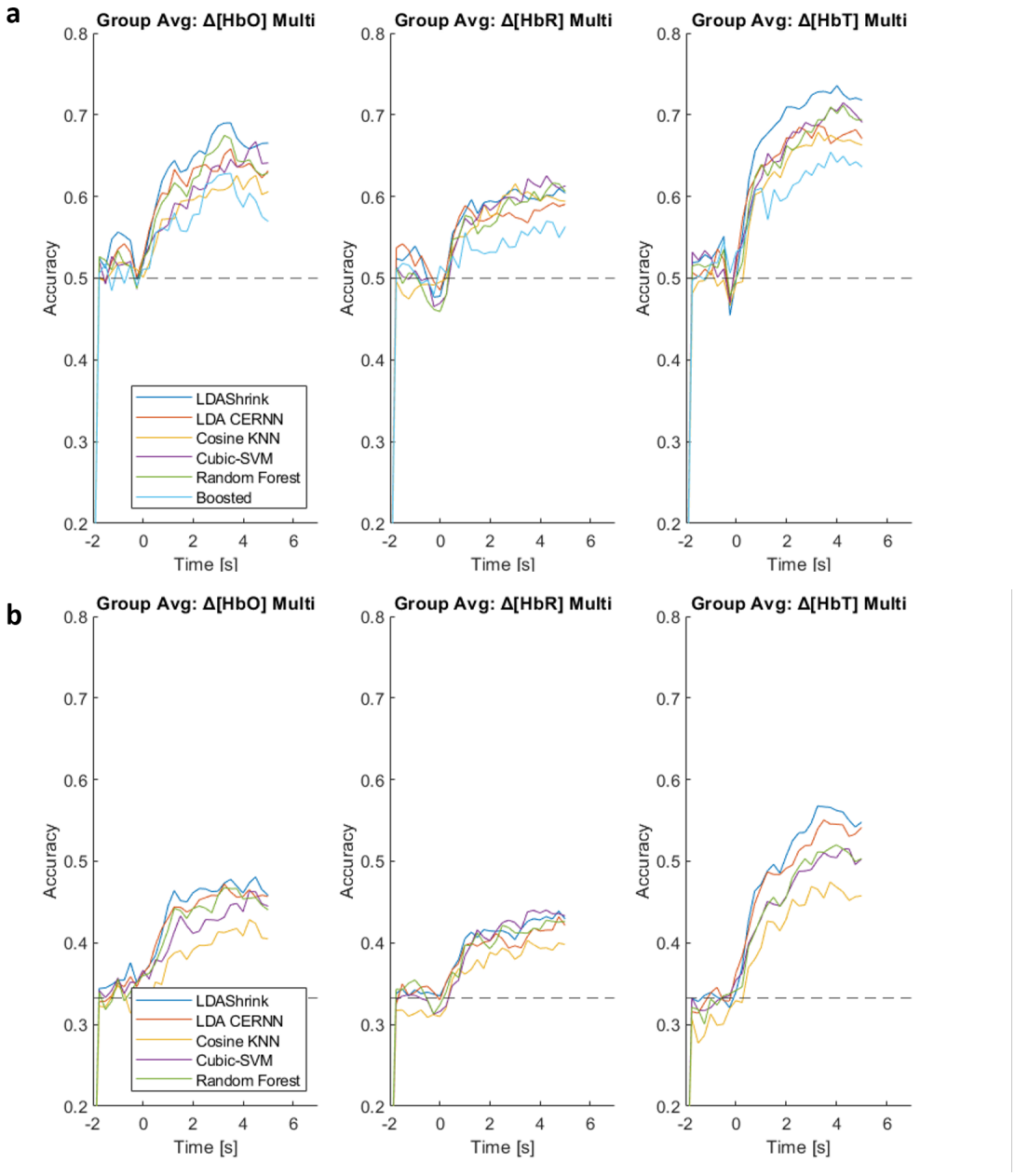

Figure 2·7: Decoding performances for other classifiers. (a) 2-Class CV accuracies for Δ[HbO] on the left panel, Δ[HbR] on the middle panel, and Δ[HbT] on the right panel. Error bars representing 95% confidence interval are omitted for clarity. (b) Same as (a) but for 3-class CV accuracies. It's very noteworthy that linear discriminant analysis with linear shrinkage has the highest classification accuracy for most of the trial segment for Δ[HbT].

tion, the cross-validation accuracy is already significantly above chance level using 1 sec window length ($t(6) = 2.94$, $p = 0.026$). In addition, the lateralization of decoding performance to the right hemisphere is evident in both FEF and IPS regions. However, FEF has substantially higher decoding performance than IPS. Even when combining left and right IPS regions for a total of 8 channels to account for FEF's higher dimension, it remains at chance level (not shown). Seeing that the R FEF region is the primary driver of decoding performance, in order to identify specific key channels within the R FEF region, we performed LOFO-classification for all channels (Fig. 2·9). All top 3 channels within the R FEF region are located closer to the medial axis (Inset of Fig. 2·10). Specifically, these channels cover FEF and superior precentral sulcus. While FEF and IPS both are part of the dorsal frontoparietal network known to play an important role in visuospatial attention, our study shows that FEF has substantially better fNIRS decoding performance than IPS.

2.4 Discussions.

In fNIRS study, we demonstrated the capability of fNIRS to decode attended spatial location in CSA. During quantitative assessment of the classifier, we made several discoveries. First, we found that Δ[HbT] provides substantially higher CV accuracy than Δ[HbO] and Δ[HbR], consistent with a previous report (Xu et al. 2013). Second, the CV accuracy of a classifier was significantly correlated with behavioral performance. It is well known that attention can modulate cortical response properties, e.g., in fMRI studies on auditory primary and secondary cortices (Jäncke, Mirzazade, and Joni Shah 1999; Petkov et al. 2004). The difference in CV accuracy between high and low performance groups supports the idea that attention improves decoding performance. Third, we identified right FEF region as making the largest contribution to overall decoding performance. Interestingly, FEF had substantially higher decoding performance compared to IPS even though both are part of dorsal frontoparietal network, which has been implicated in previous studies of spatial attention

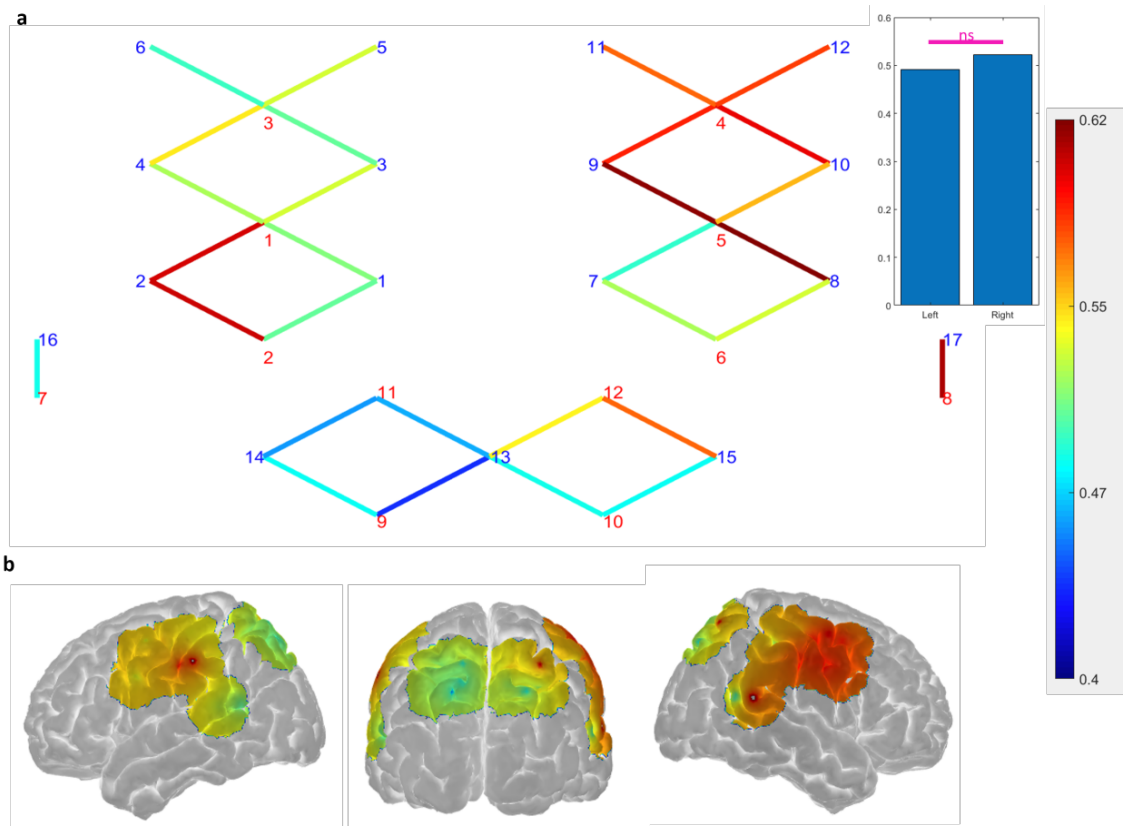

Figure 2·8: Single-Channel CV Accuracies. (a) 2-D probe layout where each line represents a channel with the corresponding colormap indicating their CV accuracies. Red numbers represent source optodes and blue numbers represent detector optodes. Inset: single chancel classification grouped by hemisphere, reported as mean and standard deviation. Left: 0.51 ± 0.047. Right: 0.56 ± 0.050. ns is not significant. Two-tailed two-sample t-test for lateralization: t(28) = -1.81, p=0.081. (b) Colormap of the single-channel CV accuracies plotted on the surface of the brain with interpolation. Mean and median of 95% CI across all channels are 0.168 and 0.135 respectively.

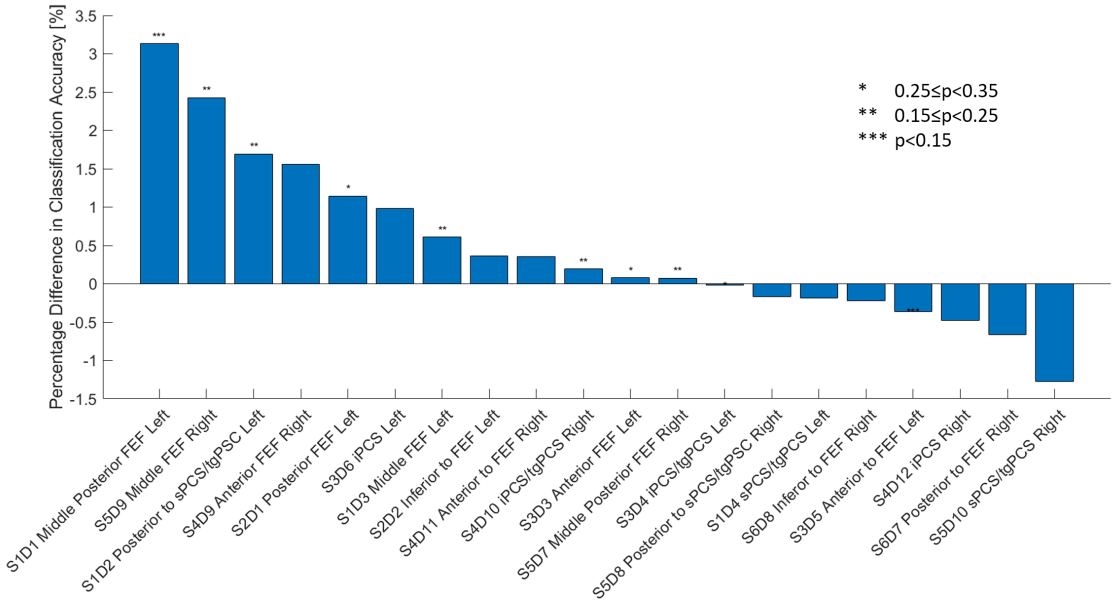

Figure 2·9: FEF has biggest impact as determined by leave-one-feature-out classification. Bar chart showing the averaged percentage difference in CV accuracies between all-channel classification and leave-one-channel-out classification.

(Shomstein 2006; Smith et al. 2010; C.-T. Wu et al. 2007). Finally, we found that CV accuracy improved faster than expected, reaching significantly (at $\alpha = 0.05$) above chance level at 1s, well before of the peak of the HRFs. This suggests that the onset of the HRFs, not just the peaks which occur much later, can contain useful information for classification. A contributing factor in the increased classification performance in high performing subjects is the lower variability of the underlying HRFs in high performing subjects (Figure 2·4). Signal detection theory predicts that higher detection and discrimination performance can result from an increase in the mean difference and/or a decrease in the variability of the two distributions (Green and Swets 2000). A previous study found reduced variability of neural responses in auditory cortex associated with improved neural detection during auditory task performance (Trapp et al. 2016). Such an effect is a potential neural mechanism underlying the HRFs in our experiments.

One advantage of fNIRS over EEG is better spatial resolution, which revealed FEF

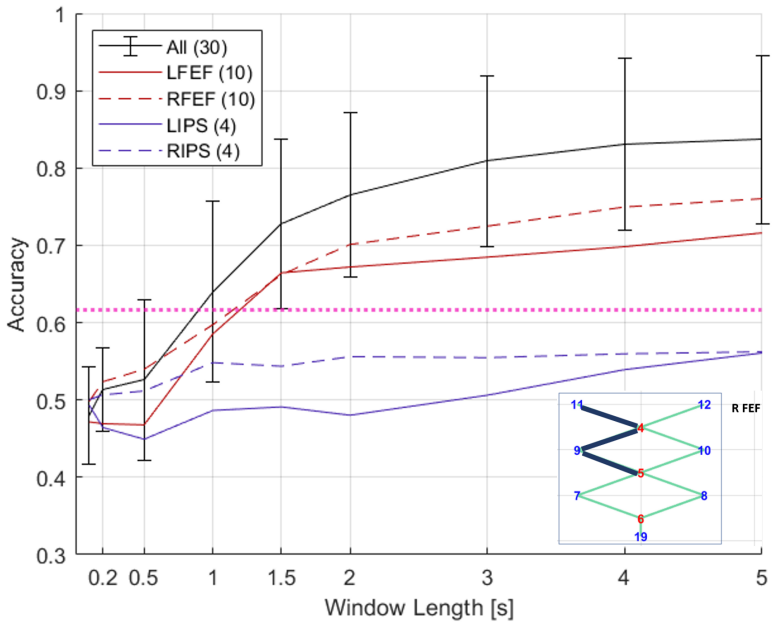

Figure 2·10: Time course of improvement in CV accuracy. Average classification accuracies of high-performing group as a function of time. All windows start at 0s (cue onset). Decision window lengths tested here are 0.1s, 0.2s, 0.5s, 1s, 1.5s, 2s, 3s, 4s, and 5s. LFEF = left frontal eye field. RFEF = right frontal eye field. LIPS = left intraparietal sulcus. RIPS = right intraparietal sulcus. The numbers next to subregion in legend represent the number of channels. Pink line indicates statistical significance for p = 0.05 using two-tailed t-test and the given standard deviation from our cross validation accuracies across subjects at window length 1s. Error bars represent 95% confidence intervals. Error bars for other lines are similar in magnitude and are not shown for visual clarity. (Inset) Top 3 channels from R FEF based on LOFO impact are shown as navy lines.

to be a critical region for decoding the attended location in CSA. Moreover, relative to EEG, fNIRS is less influenced by eye movement and blinking, enabling the use of visual stimuli for decoding brain signals without significant contributions from eye movements. In the future, it would be interesting to use an integrated fNIRS-EEG system to leverage the high spatial resolution of fNIRS and the high temporal resolution of EEG to develop a robust and rapid decoding algorithm. It also would be interesting to see whether fNIRS can be applied in assistive devices to help those struggling with CSA, e.g., populations with ADHD, autism and hearing impairment.

Chapter 3

Fast EEG-based Decoding of Attended Audio-Visual Spatial Location.

3.1 Introduction

In this chapter, we investigate the use of EEG to decode spatial attention of an attended audio-visual stimulus. While there are several successful EEG-based algorithms, this study addresses two key deficits. First, to date, most focused exclusively on the auditory modality. In complex natural scene, a more ecologically typical scenario, e.g., a cocktail party, vision also plays a critical role in directing attention to audio-visual objects, e.g., a target speaker. Thus, careful considerations must be taken in with respect to both audio-visual stimuli and ocular-related activities. In particular, one key observation about ocular-related activities and especially saccade is that the lateral differential effect of saccade is largely confined to delta and theta frequency bands (Gasser, Sroka, and Möcks 1985; Hagemann and Naumann 2001; Whitton, Lue, and Moldofsky 1978). Second, several decoding strategies have been based on reconstructing the attended auditory stimuli based on stimulus reconstruction strategies (Alickovic et al. 2019; Aroudi et al. 2019; Cheveigné et al. 2018; Mirkovic et al. 2015; Narayanan and A. Bertrand 2020; J. A. O'Sullivan et al. 2015; Taillez, Kollmeier, and Meyer 2020; Wong et al. 2018). Requiring access to the original stimulus presents an obvious problem. While some algorithms attempted to solve this with speech segregation algorithm, speech segregation algorithm by its own entity adds a lot of overhead. Another class of ADD algorithms is based on EEG signals themselves and encompasses commpon

spatial pattern (CSP) (for background on CSP, see Blankertz et al. 2008), (Geirnaert, Vandecappelle, et al. 2021) and convolutional neural network (CNN) (Ciccarelli et al. 2019; Su et al. 2022; Vandecappelle et al. 2021) and have showed strong classification accuracy. These have the benefit of being independent of stimulus features. Here, we extend CSP algorithm to decode the spatial locations of audio-visual object. Our results demonstrate robust and rapid decoding of attended spatial location of audio-visual stimuli based on the EEG signals using CSP analysis. Moreover, we find that distinct frequency bands show distinct temporal dynamics of decoding. In particular, delta and theta frequency bands show high decoding performance at cue onsets and movie offsets, which coincide with the expected timings of saccades whereas all higher frequency bands exhibit high decoding performance that is consistent throughout the duration of the trial. These findings reveal novel aspects of EEG decoding based on audio-visual stimuli, compared to auditory stimuli, suggesting important cross-modal interactions consistent with previous studies (Bizley, Nodal, et al. 2007; Bizley, Maddox, and Lee 2016).

3.2 Materials and Methods

This study employs 64Ch BrainCap MR, a 64-channel EEG-recording cap made by EASYCAP. All electrodes are sintered Ag/AgCl sensors. For the electrodes located on the cap, they are layered with 5 kOhm resistors. For the drop-down electrodes for electromyography and electrocardiography, they are layered with 15 kOhm resistors. Gel is applied topically to minimize unwanted impedance between sensor and scalp. The data were recorded in an electromagnetic-shield room at 5 kHz with FCz as the reference channel and AFz as the ground channel. The EEG-recording cap is connected to the amplifier (BrainAmp-MR, BrainProducts). Two of the three electromyographic (EMG) channels were repurposed to horizontal electro-oculographic (hEOG) channels and the 3rd EMG channel as well as electrocardiographic (ECG) channel both weren't used in the experiment.

Dataset. The videos used in the experiment are from AVSpeech, a publicly available dataset (https://looking-to-listen.github.io/avspeech/index.html). For target stimuli, each subject was shown 10 different speakers, each speaker having 5 different clips, all showing the speakers' faces clearly. Each clip was shown at all spatial locations at equal frequencies. For masker stimuli, each subject was shown 50 different speakers which are separated from the pool of speakers designated for target stimuli. These 50 speakers also have 5 different clips. For both pools of targets and maskers, all are English speakers. Each subject was presented with different pools of speakers and maskers. For each video clip, the audio clip was extracted and passed through three different head-related transfer functions (HRTF) to create auditory virtualization, one for each location (center, 45° to the left, and 45° to the right). The HRTF filters are based on measurements using KEMAR head model (https://sound.media.mit.edu/resources/KEMAR.html). For each location, the HRTF-filtered audio clip replaced the original audio waveform in the video clip. Thus, there is a total of 300 trials, 50 for each location and condition. Stimuli delivery software was written with Psychtoolbox-3 library.

Participants. 12 adults with normal hearing (age 18-44) were recruited for this study in accordance with the Institutional Review Board of Boston University. COVID-19 protocol was developed and strictly adhered to in accordance with the Institutional Review Board of Boston University. Participants were screened to exclude those with neurological disorders. Participants were briefed and consented before partaking in this study and were compensated for their time.

Experiment. The experiment is almost identical to our previous fNIRS-based experiment (Ning et al. 2022). For the calibration, before the start of the main experiment, we recorded EEG signals while the subjects were instructed to sit still and only move and blink their eyes in the following order: left, right, up, down and blink 5 times, 10 secs between. A computer with three monitors was used for the experiment. The monitors were located

at the three locations: center, 45° to the left, and 45° to the right (Fig. 3·1). They all are equidistant from the center where the subject is sitting upright. The auditory stimuli were delivered via earphone (ER-1 Etymotic Research Inc.) with eartips (E-A-RLink 3A Insert Eartips). This study adopted event-related design. For each trial, a 2 seconds long audio-visual cue was delivered in the form of a white cross against a black background and a 2 kHz pure tone linearly ramped in the first 0.5 s. This audio-visual cue appeared randomly at one of the 3 spatial locations, indicating the location of the target speaker. For the target-alone (TA) condition (Fig. 3·1a), the cue was followed by a video of the target speaker. For the target-plus-masker (TM) condition (Fig. 3·1b), the cue was followed by 3 videos, one for each location, one of which is the target speaker and the 2 remaining are the maskers. Afterward, in both conditions, it's followed by two multiple-choice questions, each question containing 5 possible choices. These multiple-choice questions are always displayed at the center location. The two questions are related to face identification and words identification, presented in that order. In the face identification task, the subjects are presented five different faces and is tasked with correctly identifying the face of the target speaker shown in the video. In the word identification task, the subjects are presented five different transcripts and is tasked with correctly identifying the words spoken by the target speaker in the video. Thereafter, an instruction to press the space bar to begin the next trial is displayed at the center location. While the audio-visual cue and the video clip last 2 and 3 seconds respectively, the subjects have 20 seconds in total to answer both multiple-choice questions but can move on to the space bar instruction immediately after the questions are done. The subjects are given a 2 to 4-minute-long break in the middle of the experiment. Lastly, the subjects are provided with chin rest in order to discourage head movements.

Grouping by Task Performance. Using the same criteria in our previous fNIRS study (Ning et al. 2022), we divided subjects into 2 groups, high-performing group where subjects scored at least 80% correct on all trials, and low-performing group where subjects scored

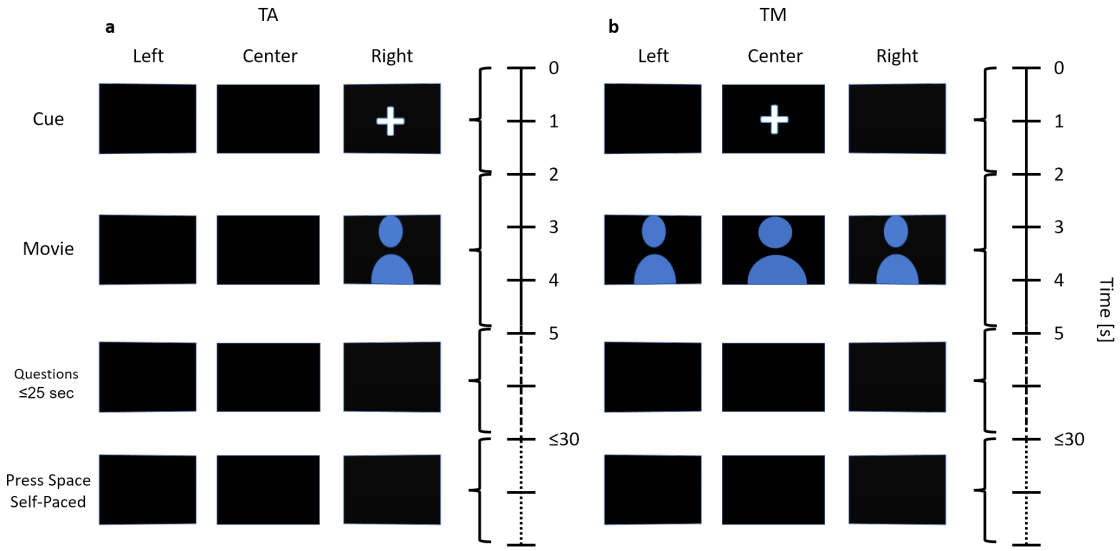

Figure 3·1: Schematic Diagram of the Trials. (a) Timeline of the trial for TA condition. Each row represents the three monitors at a point in time. A cue in the form of a white cross accompanied by a pure ramping tone of 2 kHz randomly appears at one of the three locations to indicate the location of the target stimuli at the start of the trial. In this example, the cued location is at the right. Next, a video clip is displayed simultaneously for 3 seconds. Next, two multiple-choice questions are displayed at the center monitor, to be answered with keyboard. 1st question is to identify the face of the speaker, 2nd question is to identify the words spoken by the target speaker. Subject has up to 25 seconds to answer both questions. Upon the completion of the questions, instruction to press the space bar is displayed at the center monitor to begin the next trial. (b) Timeline of the trial for TM condition. It's identical to TA condition except three video clips are displayed simultaneously instead of one. In this example, the cued location is the center.

less than 80%. A trial is counted as correct if the participant correctly identified both the faces of speakers and the words spoken by the targeted speakers.

EEG Preprocessing. For the delta to gamma-low frequency bands, EEG signals were low-pass filtered with passband edge frequency of 40 Hz (with cutoff frequency 45 Hz at -6 dB and 10 Hz transition bandwidth) using zero-phase, non-causal Hamming window filter of order 1650 before down-sampled to 250 Hz. Next, they were high-pass filtered with passband edge frequency of 1.5 Hz (with cutoff frequency 0.75 Hz at -6 dB and 1.5 Hz transition bandwidth) using zero-phase, non-causal Hamming window filter of order 550. Next, channels with standard deviation above 75 mV were automatically rejected. Trials with variance above the threshold (defined as 80th-percentile + 3*abs(80th-percentile-20th-percentile)) were rejected. Next, we opted for interference subtraction method using linear least square method to reduce ocular artifacts (Parra et al. 2005). This will be called EOG regression. Next, we band-passed the signals into 6 different frequency bands where the frequency parameterized are the passband edge frequencies: delta band (1.5-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), low-beta (13-17 Hz), high-beta (18-30 Hz), and gamma-low (30-40 Hz). No common-average referencing was performed because some electrodes have significantly higher impedance or line noise (60 Hz) than other electrodes even after filtering and channel-rejection, resulting in unbalanced weight averaging.

While the gamma-low range used here is quite narrow, we tested the range 30-50 Hz and found no substantial difference in decoding performance (not shown). The narrow range is necessary for reducing noise associated with power transmission line (60 Hz).

For the gamma-high frequency band, EEG signals were low-pass filtered at 150 Hz (with cutoff frequency 168.75 Hz at -6 dB and 37.5 Hz transition bandwidth) using zero-phase, non-causal Hamming window filter of order 440 before down-sampled to 500 Hz and then high-pass filtered at 100 Hz (with cutoff frequency 87.5 Hz at -6 dB and 25 Hz transition bandwidth) using zero-phase, non-causal Hamming window filter. Channels with

standard deviation above 125 mV were rejected. Thereafter, the preprocessing pipeline is identical for the delta to gamma-low bands.

EEG Classification. For each frequency band, we performed common spatial pattern (CSP) filtering on 1 sec window for both 2-class and 3-class classification. We begin the cross-validation step here by dividing the EEG trials into training and test sets. We extracted 6 common spatial pattern (CSP) filters from the training set for each frequency band, 3 filters with the maximum ratio of variances for each condition. These 6 filters are applied on the test set. Next, the features are extracted as the logarithm of the variance of the CSP-filtered signals, where the variance represents variation over time. In addition, we also tested filter-bank CSP where we combine 6 CSP filters from each of the following frequency bands (theta, alpha, low-beta and gamma-low) to form a total of 24 filters. For the 3-class classification, before extracting CSP filters from the training set, we added a small constant (σ = 1e-8) to the diagonal elements of covariance matrices before performing joint-approximate diagonalization to ensure that the covariance matrices are well-conditioned, a process commonly known in statistical signal processing literature as diagonal loading (Carlson 1988).

In this study, to test the cross-validation (CV) accuracy of the classifier trained with the entire probe, we tested 2-class classification between left (-45°) and right (45°) spatial locations as well as 3-class classification between all 3 spatial locations. Linear discriminant analysis with covariance matrix estimator based on linear combination between sample covariance matrix and identity matrix is used (Ledoit and Wolf 2004). The one-sec window is tested for every interval starting from 2.5 sec before the start of the cue to 7.5 sec after the start of the cue, in step of 1/5th sec.

3.3 CSP

CSP finds a linear transformation of two covariance matrices that maximizes the ratio of the variances between two classes of signals for discrimination. Consider here the case of binary classification. Roughly, CSP can be seen as the supervised version of principal component analysis (PCA) by changing from the singular value decomposition for the overall covariance matrix to the generalized eigenvalue problem (GEP) for a pair of covariance matrices (A,B), one for each class (Blankertz et al. 2008). The full derivation of CSP is left in Chapter 5.

3.4 Results

Behavioral Results. All 12 subjects performed significantly above chance level in both tasks (mean ± standard deviation of all trials: 84.3±14.1%). Furthermore, all 12 subjects performed above 80% on TA trials whereas 4 out of 12 subjects performed less than 80% on TM trials. Of these 8 subjects, who had performance levels of 80% or above on TM trials were categorized as high-performing, and 4 subjects with performance levels less than 80% on TM trials were categorized as low performing. One subject performed 0% in left vs right for the TM condition so that subject was excluded from all TM analysis. Thus, we have 12 subjects for analysis in the TA condition and 11 subjects for TM condition.

Classification Performance of the Filter-Bank. The grand averages and their 95% confidence intervals of filter-bank (which include theta, alpha, low-beta and gamma-low frequency bands) CSP 2-class decoding performance as a function of the ending time of the 1-second window are shown in the top row of Figure 3·2. The performance first crossed the 70% accuracy at 900ms after the cue onset in the TA condition and remains significantly above chance level until after movie offset. Similarly, the performance first crossed the 70% accuracy at 500ms after the cue onset in the TM condition. Next, we split the subjects into two groups based on their task performance. The averages of high-performance group

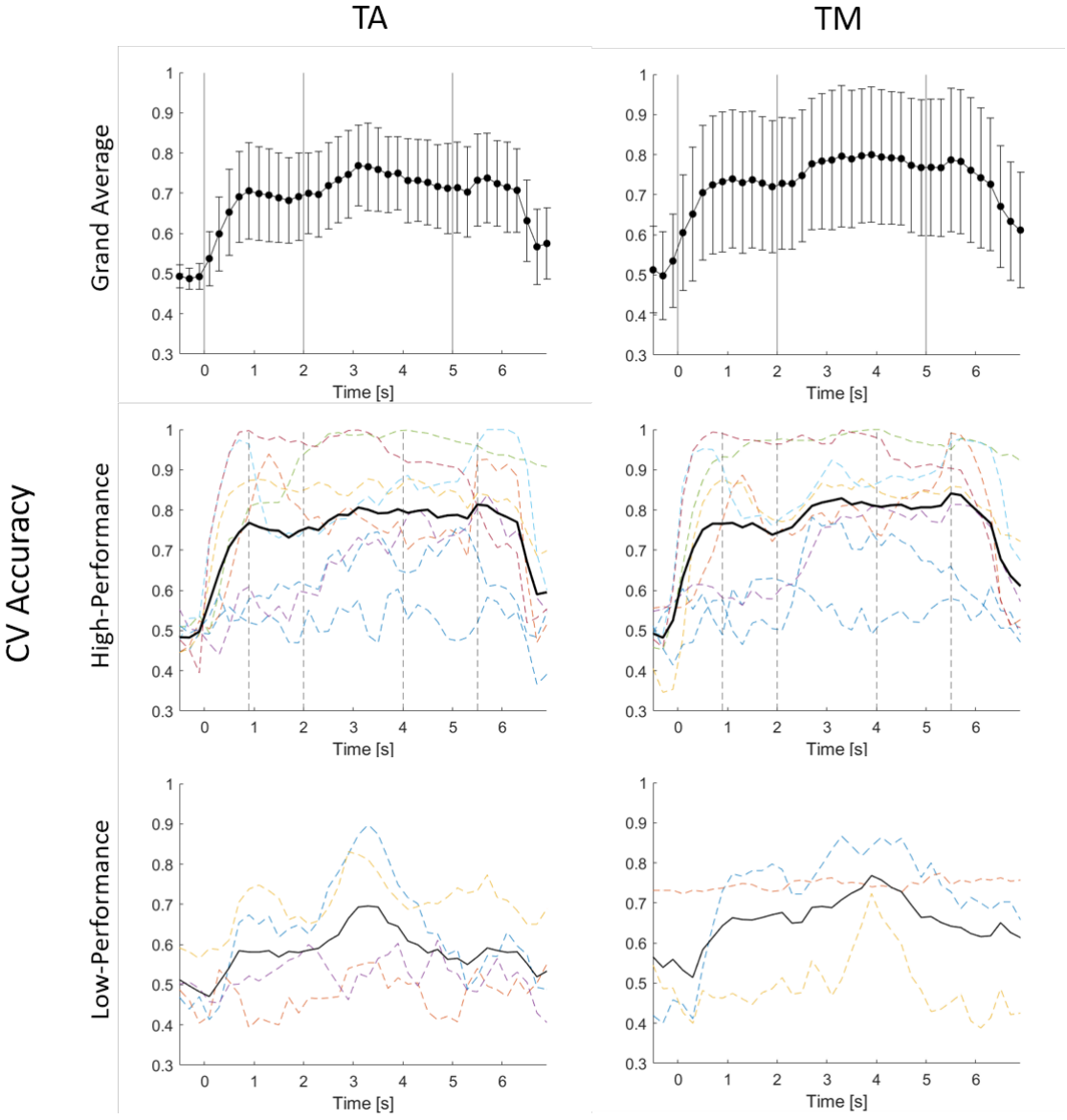

Figure 3·2: Decoding Performance of the Filter-Bank. Decoding performance of the filter-bank (theta, alpha, low-beta and gamma-low) as a function of the ending time of a decision window 1 s long. Top row: grand averages and their 95% confidence intervals. In chronological order, vertical lines represent cue onset, movie onset and movie offset. Middle row: averages of high-performing subjects in bold black line. Individual subjects shown as colored dashed lines. The dashed vertical lines in the middle row represent the timings used in Figure 3·5. Bottom row: averages of low-performing subjects in bold black line. Individual subjects shown as colored dashed lines. Left column: TA condition. Right column: TM condition.

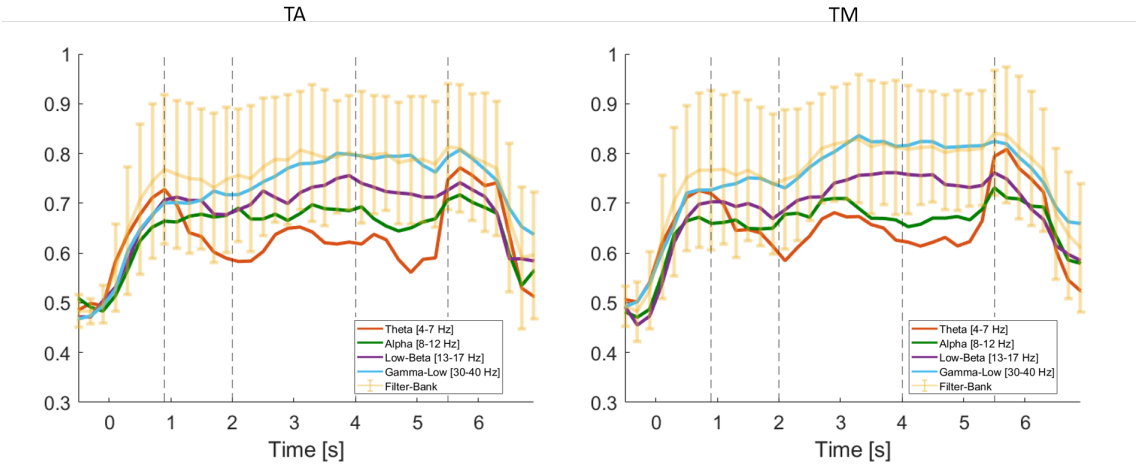

Figure 3·3: Band-Specific Decoding Performance. Similar to Fig. 3·2, decoding performance for high-performance group as a function of ending time of decision window 1s long. The 4 frequency bands shown are theta band, alpha band, low-beta band and gamma-low band. Decoding performance for TA condition is shown in left and TM condition in right.

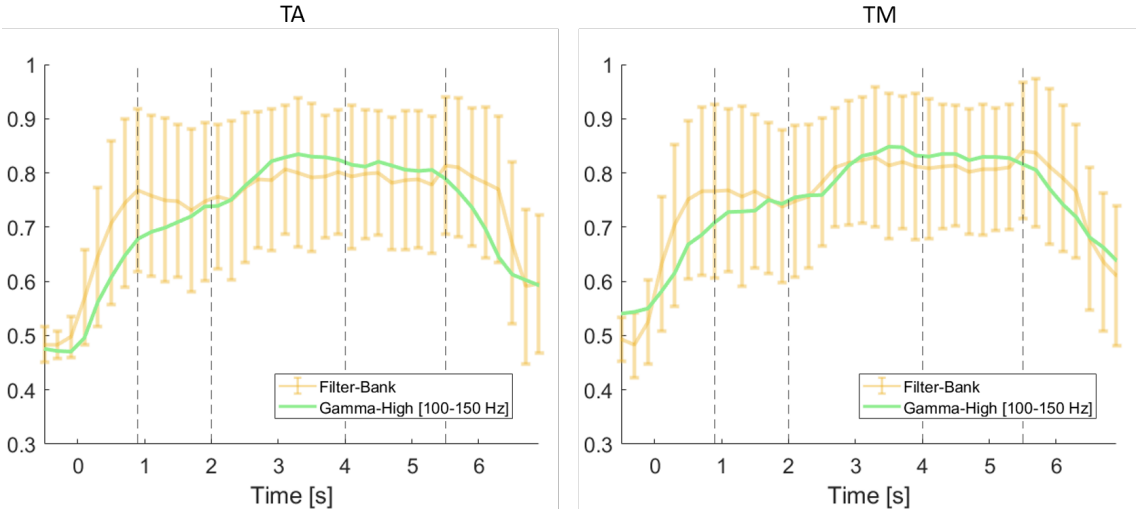

Figure 3·4: Decoding Performance Based on Gamma-High. Similar to Fig. 3·3, decoding based on gamma-high [100-150 Hz] band and filter-bank. Note that filter-bank does not include gamma-high frequency band.

and low-performance group are shown in the middle and bottom row of Figure 3·2, respectively. In addition, the average decoding performances for each individual subjects are shown as dashed colored lines in middle and bottom row of Figure 3·2. This is to highlight the variability in decoding performance across subjects. The difference in decoding performance between high and low-performance group supports the hypothesis that the decoding performance is highly dependent on task-performance.

Band-Specific Performance. The averages of CSP decoding performances of high-performance group for the 6 frequency bands as well as the filter-bank are also shown in Figure 3·3. All combinations of frequency bands and conditions (TA vs TM) yield above chance-level classification accuracy shortly after the onset. The bimodal peaks of the theta band (red lines in Figure 3·3) in both conditions reflect the expected timings of lateral eye movements. On the other hand, the decoding performances based on the gamma-low band lack the bimodal peaks and is more sustained throughout the trial (light-blue lines in Figure 3·3).

We separately assess the decoding performance based on the gamma-high [100-150 Hz] frequency band (Fig. 3·4) which is unlikely to originate from the brain. The decoding performance of the gamma-high band rises slowly relatively to the filter-bank and eventually bypasses the performance of the filter-bank starting at the movie onset. This analysis suggests that decoding the spatial location is also possible using physiological signals that do not originate from the brain.

We noted two trends in Figure 3·3. First, the decoding performance increases as the frequency band increases for t = 1-2s and 3-4s, which is when participants are expected to have fixed gaze at target spatial locations. Second, the timings of the bimodal peaks of theta frequency bands suggest that these peaks occur approximately around the time of the eye movements. Thus, we performed one-way repeated measure ANOVA tests to test across two different groupings. First, we look at whether the difference in decoding performance

	F(3,21)	p-value	
Fixed Time (TA)			
-0.1-0.9s	0.67	0.51	
1-2s	2.15	0.18	
3-4s	8.48	0.0057	**
4.5-5.5s	1.74	0.20	
Fixed Time (TM)			
-0.1-0.9s	0.88	0.42	
1-2s	2.88	0.12	
3-4s	8.02	0.0066	**
4.5-5.5s	4.31	0.032	*
Fixed Frequency (TA)			
theta	3.79	0.035	*
alpha	0.36	0.77	
low-beta	0.88	0.42	
gamma-low	2.62	0.13	
Fixed Frequency (TM)			
theta	6.79	0.015	*
alpha	1.66	0.22	
low-beta	2.47	0.12	
gamma-low	2.82	0.12	

Table 3.1: rANOVA Results. The rANOVA results were Greenhouse-Geisser corrected for non-sphericity using the epsilon coefficient. In all cases, the degree of freedom are 3 for the factor tested and 21 for the error term. Note that although the overall rANOVA result for the fixed theta frequency-band in TA condition is significant, none of the pairwise t-test results are significant as shown in Figure 3·6

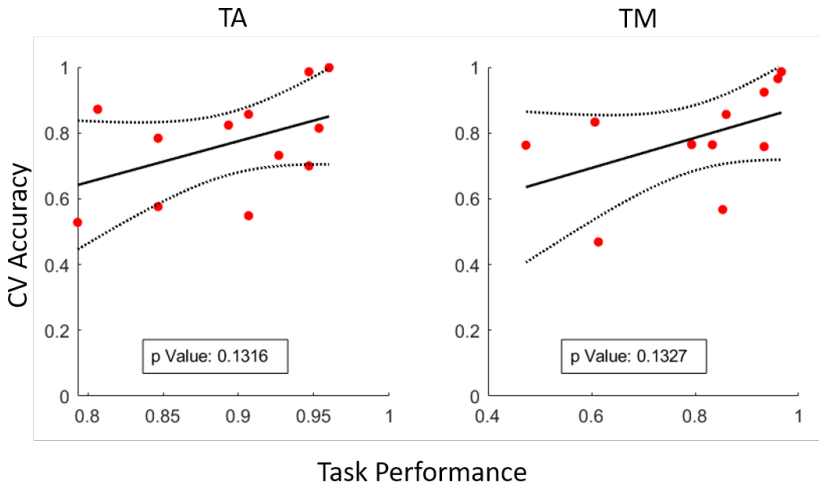

Figure 3·5: Correlation Between Filter-Bank Decoding Performance and Task Performance. Linear regression between task performance and decoding performance using window 2.1-3.1s for both TA and TM conditions. For TA condition, n = 12. For TM condition, n = 11. Dashed lines represent 95% confidence intervals. Left is for TA condition and right is for TM condition.

across different frequency bands within a fixed time point is statistically significant. Next, we look at whether the difference in decoding performance across different time points within a fixed frequency band is statistically significant. This is run once for each condition (TA and TM). Thus there is a total of 16 rANOVA tests. The results for the first grouping are presented in the top row of Fig. 3·6 , and similarly, the second grouping are presented in the bottom row. For the windows 1-2s and 3-4s, which are during the cue presentation and stimuli presentation, respectively, the decoding performance increases as the frequency band increase. The difference is statistically significant for 3-4s window in both TA and TM condition (top row of Fig. 3·6). In contrast, the difference in decoding performance across different time point within a fixed frequency band is only statistically significant in theta for the TM condition. This, in conjunction with the timing of the expected eye movement as shown in Fig. 3·3, suggests that the decoding performance of theta frequency band is highly sensitive to the eye movement. The F-statistic and Greenhouse-Geisser corrected

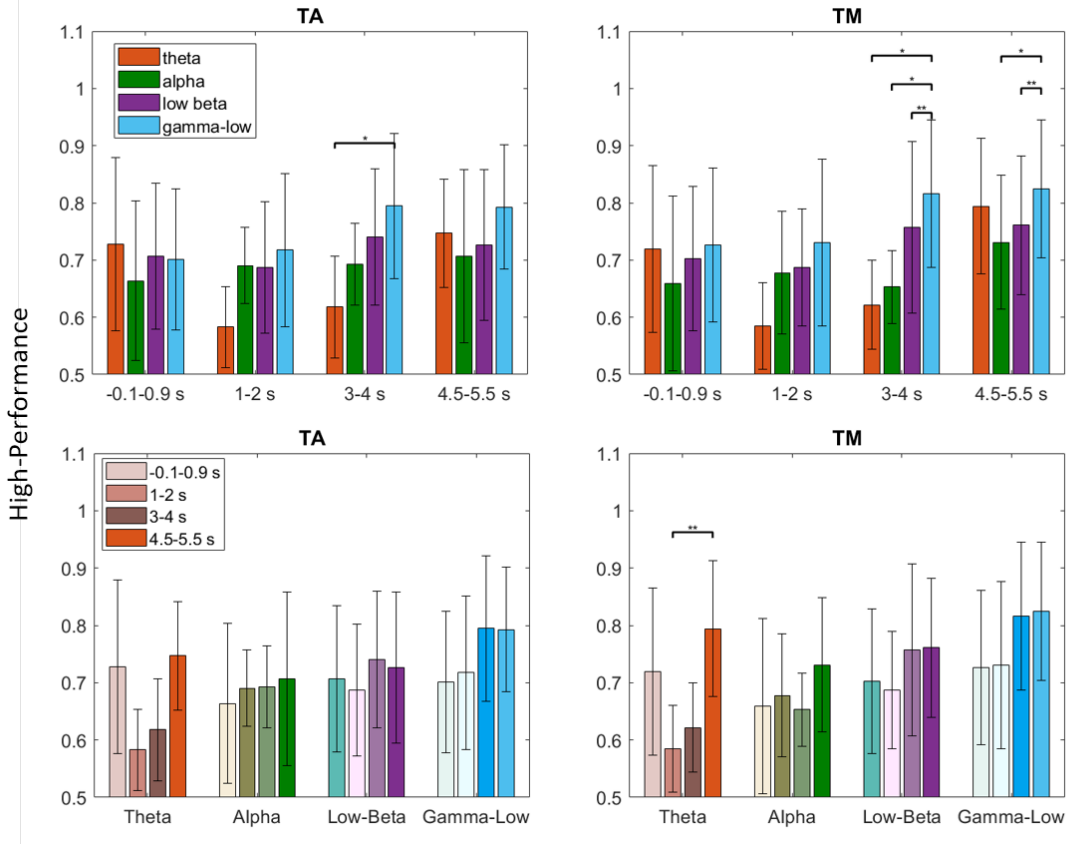

Figure 3·6: Repeated-Measures ANOVA Tests. Top row tested difference in decoding performance across different frequency bands within a time point. Bottom row tested different in decoding performance across 4 different time points within a frequency band. 16 repeated-measure ANOVA tests are conducted in total, one for each group. At t=1-2s and t=3-4s, in both TA and TM conditions, the decoding performance increase as the frequency band increase. In the bottom right panel, for the theta band in TM condition, the decoding performances peaked at the onset of the cue and the offset of the movie, in accordance with the expected timing of lateral eye movements. All bars represent averages of all high-performance subjects as defined in the main text. Error bars represent 95% confidence interval across subjects. p-values are reported in Table 3.1.

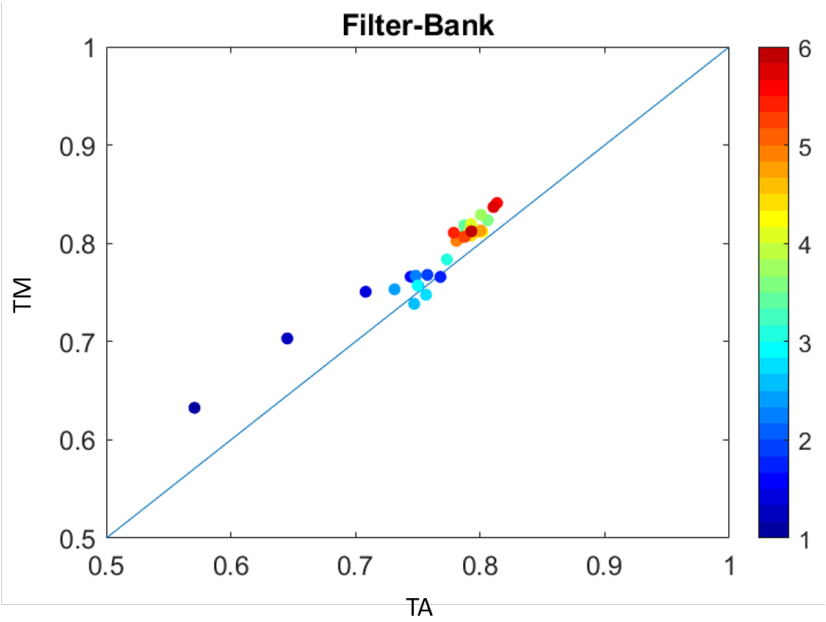

Figure 3·7: Scatter Plot of Filter-Bank Decoding Performance between TA and TM Conditions for High-Performance Group. Scatter plot between TA and TM conditions for high-performance group using filter-bank (theta, alpha, low-beta and gamma-low) CSP algorithm. Different colors represent ending times of the decision window 1s long.

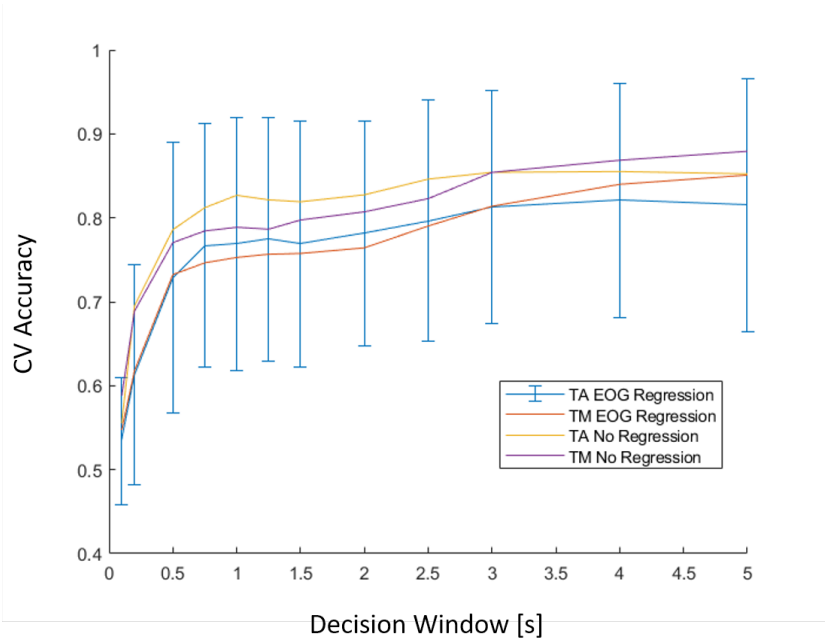

Figure 3·8: Decoding Performance as a Function of Window Lengths. Decoding performance as a function of window lengths for 4 different scenarios are shown. EOG regression negatively affected the decoding performance by roughly 5%. Error bars represent 95% confidence interval. Error bars for 3 other lines are similar in magnitude and omitted for visual clarity.

p-value are displayed in Table 3.1.

Task Performance Correlation. Next, we fitted the linear regression between CV accuracy of the filter bank and task performance (Figure 3·5). For both conditions, the slope coefficients of the linear regression are not statistically significant at $\alpha = 0.05$. It's important to note that the small sample size still renders it difficult to draw any definite conclusion.

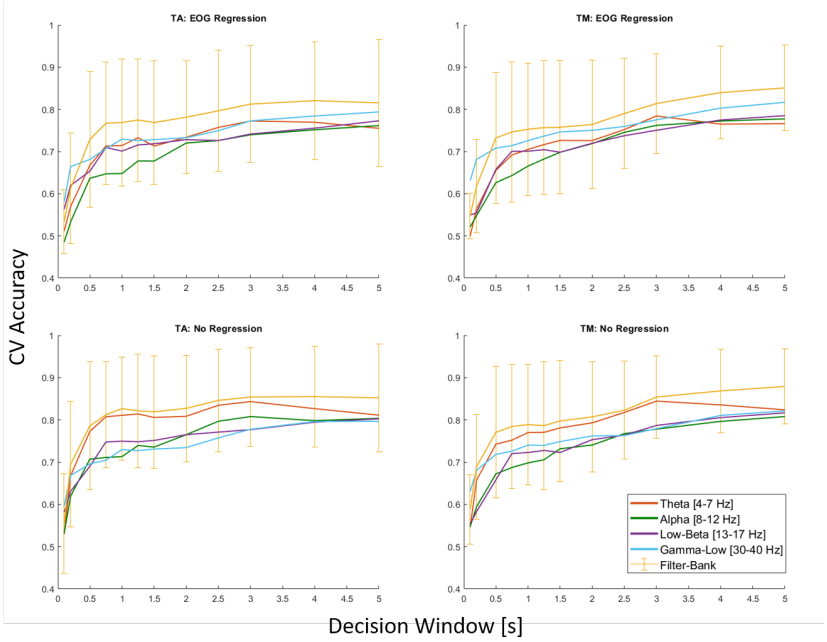

Figure 3·9: Band-Specific Latency Analysis. Same as Figure 3·8 but for 4 individual frequency bands (theta to gamma-low).

To examine the effect the EOG regression has on the decoding performance for the latency analysis, we plotted scatter plots of decoding performances between preprocessing pipelines with and without EOG regression for the 5 frequency bands using data points from Fig. 3·9 (Fig. 3·10 for TA condition and Fig. 3·11 for TM condition). We observed that the theta frequency band is sensitive to EOG regression (several sample points are above the x=y line) and the sensitivity to EOG regression increases as the frequency band increase. This is confirmed by the lack of significant difference between EOG regression

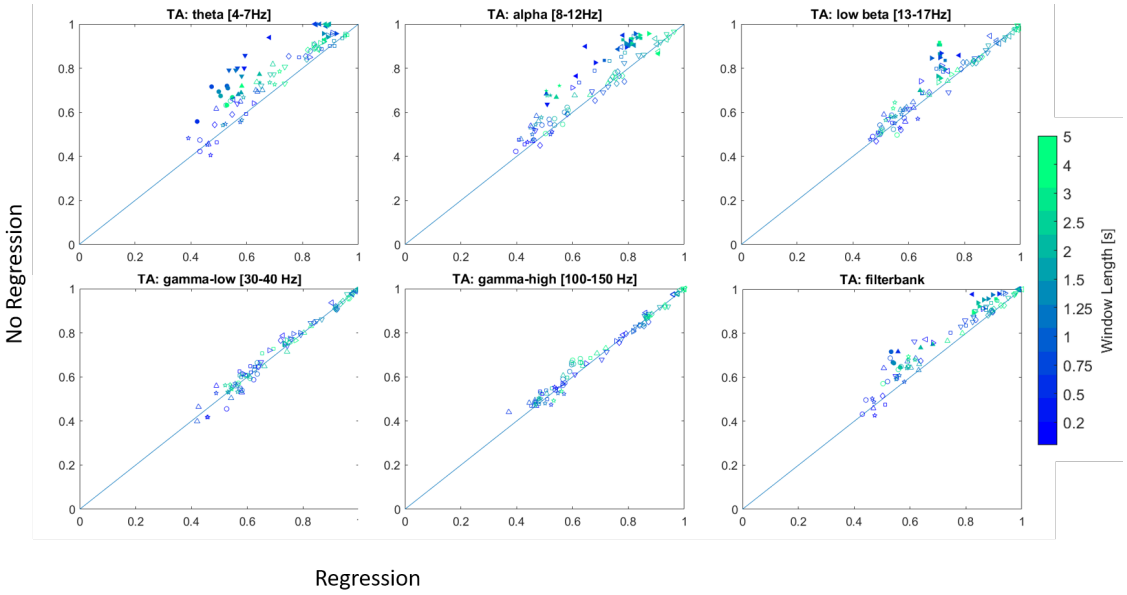

Figure 3·10: Scatter Plots between Decoding Performance with and without EOG Regression. Scatter plots of decoding performance with EOG regression at the x-axis and without EOG regression at the y-axis are shown for the 5 frequency bands as well as the filter-bank. Theta band benefits the most from ocular activities whereas gamma-low and gamma-high bands are relatively intact. Different marker shapes represent different subjects. Different color represents different window lengths, all start from cue onset. Filled marker indicates that the difference between decoding performances with and without EOG regression is statistically significant for significance level p = 0.05. All are for TA condition. All are high-performing subjects as defined in the main text.

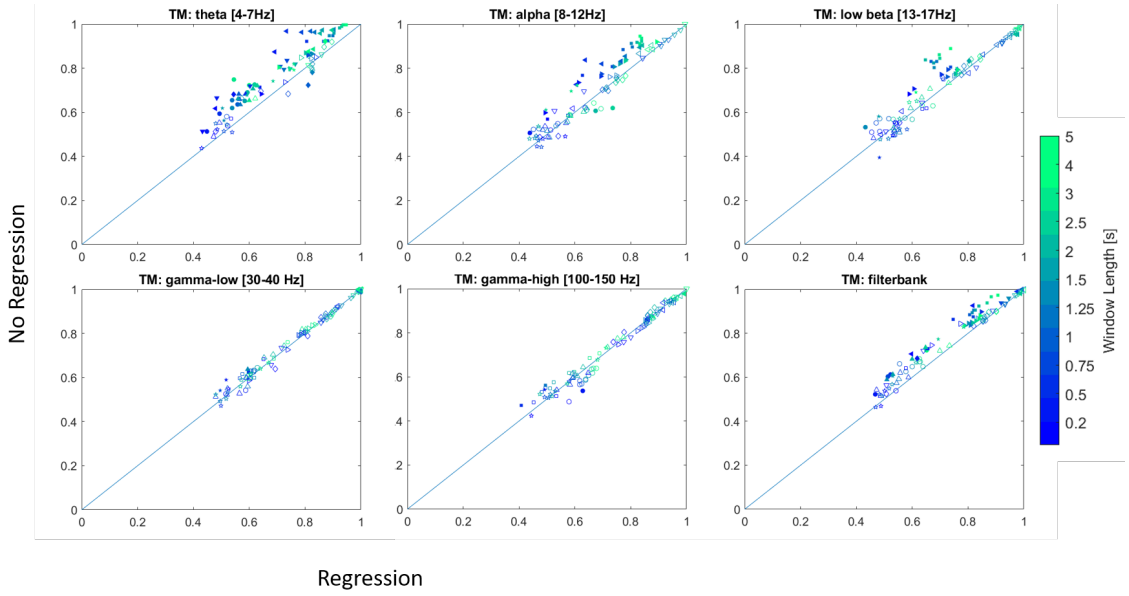

Figure 3·11: Scatter Plots for TM Condition. Same as Figure 3·10 but for TM condition.

and without EOG regression in the gamma-low and gamma-high frequency bands in both TA and TM conditions.

Maskers Effects. Next, we examine the way maskers affect both task performance and decoding performance. The decrease in task performance from TA (89.4%) to TM condition (75.9%) for all subjects is statistically significant (two-tailed paired t-test: p=0.04, t(11)=2.33). Figure 3·7 suggests that the decoding performance of TM condition is slightly higher than TA condition. However, the change in decoding performance at t=3.1-4.1s between TA and TM condition is not statistically significant for high-performing subjects (79.3% for TA condition vs 80.8% for TM condition, two-tailed paired t-test: p=0.275, t((7)=-1.19). Put together, these suggests that higher cognitive demanded from the TM condition yields marginally better decoding performance.

Latency Analysis. Figure 3·8 shows the CV accuracies of filter-bank as a function of the window lengths, where all window starts at the cue onset. We tested all combinations of masker conditions and inclusion-vs-omission of ocular-removal algorithm in the

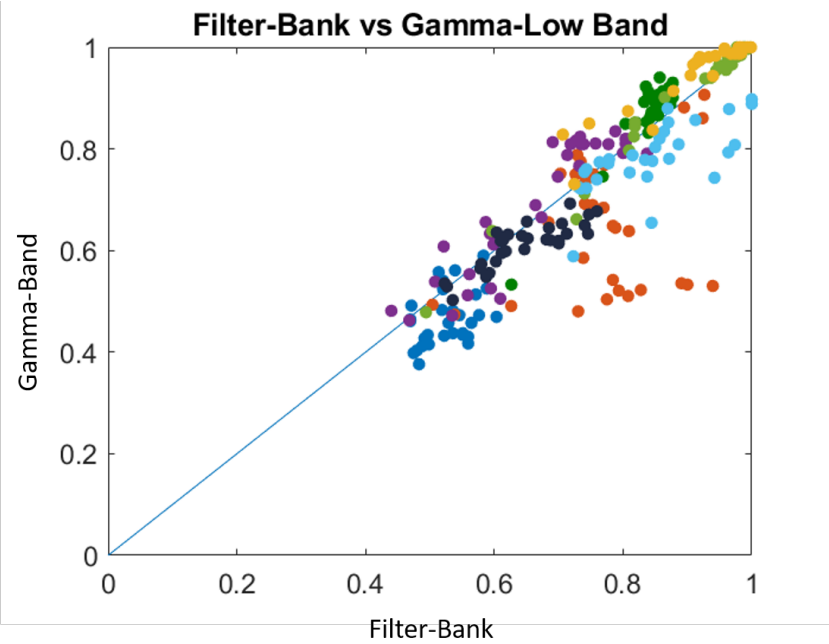

Figure 3·12: Scatter Plot between Filter-Bank and Gamma-Low Decoding Performance. Decoding performances using filter-bank method are on the x-axis whereas decoding performances trained on gamma-low frequency-band are on the y-axis. Different points within the same color represent different time points of the trial. In general, the filter-bank method substantially outperforms gamma-low band but in some subjects, gamma-low band slightly outperforms the filter-bank method. Different colors represent different subjects. All are high-performing subjects as defined in the main text.

preprocessing pipeline. We adopted EOG linear regression for ocular-removal algorithm. While the decoding algorithm benefits greatly from the omission of EOG regression in the lower-frequency range, the performance of the decoding algorithm with EOG regression still performs well and only decreased by roughly 5%. Regardless of the algorithm used, the decoding performances are relatively fast and reach significantly above chance level at 500 ms after cue onset. Since the decision windows at least 2 seconds long likely include both saccade and fixed gaze at different lateral points, this showed that common spatial pattern can be trained to both scenarios. However, when looking at individual frequency bands in Figure 3·9, an interesting pattern emerges. When including the ocular-removal algorithm, in the top row of Figure 3·9, the decoding performance of the gamma-low band outperforms the filter-bank and every other frequency band at decision windows $\leq$ 200 ms long. When excluding the EOG regression, in the bottom row of Figure 3·9, the decoding performance of the gamma-low band equates that of the filter-bank (and still outperform every other frequency band) for decision windows $\leq$ 200 ms.

Gamma Band vs Filter-Bank. A scatter-plot of the decoding performance for the TM condition between filter-bank and gamma-low frequency-band for each subject is shown in Figure 3·12. While the filter-bank outperforms the gamma-low frequency band on average, gamma-low band does slightly outperform the filter-bank in some subjects.

3.5 Discussion

3.5.1 Attention.

Consistent with previous studies showing that spatial attention in both auditory and visual modalities can lateralize the neural responses over the brain in a frequency-dependent manner, we demonstrated robust decoding of the spatial location of attention. As mentioned in the introduction, this has the benefit of not being contingent on both the availability of original speech stimuli as well as speech segregation algorithm.

Two interesting patterns can be deduced from the dynamic of performance of theta frequency bands in Figure 3·3. First, in TA condition, the decoding performance of theta increases faster (earlier) than higher frequency bands whereas in TM condition, the decoding performance of theta and gamma-low increases faster (earlier) than alpha and low-beta frequency bands. Second, the timings of the bimodal peaks of theta and delta frequency bands coincide with the expected timings of the saccades, even with EOG regression. In contrast, frequency bands above theta exhibit strong and consistent decoding performance throughout the entire duration of the trial. Put together, these reveal distinct temporal dynamics corresponding to different neural oscillations.

One way of studying attention is to divide it into fast, automatic stimulus-driven process (also known as bottom-up attention) and a slower, effortful and goal-driven process (also known as top-down attention) (Katsuki and Constantinidis 2014; Weichselgartner and Sperling 1987). In particular, neurophysiological studies from (Chalk et al. 2010; Richter et al. 2017) show that alpha-beta oscillations invoked by goal-driven attention can modulate stimulus-driven gamma oscillation. However, EEG's poor spatial resolution make interpreting coherence between different regions difficult.

Thus, hybrid EEG-fNIRS holds promise for incorporating the high spatial resolution of fNIRS to identify specific regions activated during tasks and high temporal resolution of EEG to identify neural oscillations during tasks. We can use hybrid EEG-fNIRS to identify the regions activated during different stages of cognitive tasks in order to elucidate the different types of attention engaged. In addition, we can study the synchronization/desynchronization between two or more regions using neural oscillations. In the case of audio-visual spatial attention, it can be used to explore the relationship between gamma neural oscillation and FEF and possibly other regions during the orienting of spatial attention. Since FEF is a source of top-down goal-driven attention (Petersen and Posner 2012), it can exert influence on lower bottom-up sensory-processing regions such as primary visual

cortex and extrastriate areas during visual processing via neural synchronization (Womelsdorf, Schoffelen, et al. 2007) as well as primary auditory cortex (Garg, Schwartz, and Stevens 2007; Sun et al. 2022).

One previous study utilized audio-visual stimuli to assess the decoding of spatial attention A. E. O'Sullivan, Lim, and Lalor 2019. There are two major differences between this current study and O'Sullivan's study. One difference is that the participants in O'Sullivan's study didn't allow eye movements. Our study demonstrates that robust decoding is also possible in the presence of eye movements, which would occur in more naturalistic scenes. Another difference is that O'Sullivan's study didn't directly decode the spatial location of attention. Rather, they used stimulus-reconstruction approach, which is unsuitable for real-time BCI.

3.5.2 Limitations and Future Directions.

One limitation of our study was that we did not perform eye-tracking or EMG recordings. Although, we did assess decoding performance with and without regressing out ocular-related activities, it is difficult to rule out possible contributions of other physiological signals, e.g., eye or muscle activities. The decoding performance based on high-gamma suggests that decoding may also be possible based on these physiological signals. Future studies should better dissect the contributions of physiological signals that originate outside of the brain. From a BCI decoding perspective, a combination of all of these signals can be useful for decoding the attended spatial location.

Chapter 4

Conclusions and Future Directions

4.1 CSA-Deficit Population

It should be noted that all participants in our sample for both EEG and fNIRS experiments are normal-hearing. Thus, it's necessary to perform additional experiments on people who struggle with CSA such as people with hearing loss, autism spectrum disorder and ADHD. It will be interesting to see how the responses change in frontal eye field, intraparietal sulcus, superior temporal gyrus/planum temporale and other relevant brain regions for each subpopulation (hearing loss vs autism vs ADHD). One expected difficulty to this study is the large variation due to the cortical diversity within the hearing loss population. For example, people who were born and remained deaf into adulthood will exhibit drastically different cortical structure and functioning from people who gradually lost their hearing during late adulthood. In addition, the brain is known for remarkably adaptation, also known as brain plasticity. In particular, hearing deprivation will induce cortical changes, and partial restoration of hearing with hearing aids or cochlear implant will induce additional cortical changes. For a review, see (Stropahl, Chen, and Debener 2017). Thus, it will be necessary to impose additional control on hearing loss population, e.g. regular cochlear implant users.

In addition, testing these subpopulations may also answer another question: what are the necessary and sufficient conditions for achieving fast and reliable decoding performance? Specifically, if people with hearing loss pay close attention to the stimuli (top-down attention), successfully orient their eyes to the spatial location but fail to process the

sound due to hearing loss (bottom-up), is it sufficient to achieve fast and reliable decoding performance? Similarly, what would the results look like for people with autism spectrum disorder and ADHD?

4.2 Hybrid EEG-fNIRS

As discussed in the Discussion section of EEG chapter, hybrid EEG-fNIRS holds promise for incorporating the high spatial resolution of fNIRS to identify specific regions activated during tasks and high temporal resolution of EEG to identify neural oscillations synchronized or desynchronized during tasks. Although hybrid EEG-fNIRS is still relatively new, there are already over a hundred studies as of 2021 (Z. Liu et al. 2021). For example, hybrid EEG-fNIRS has already been used to investigate the differences in activated regions during auditory or visual task (Chen, Sandmann, et al. 2015). In addition, hybrid EEG-fNIRS was also used to compare auditory and visual responses between normal-hearing with cochlear-implant users (Chen, Stropahl, et al. 2017). When used as BCI, hybrid EEG-fNIRS have already been used to successfully decode the perception as either auditory or visual (Putze et al. 2014). In addition, there are already at least two companies selling commercial hybrid EEG-fNIRS device (NIRx Medical Technologies, LLC, Berlin, Germany and Artinis Medical Systems, Einsteinweg, The Netherlands). Since it's still relatively new, as highlighted in (Z. Liu et al. 2021), one main challenge is the successful integration of EEG and fNIRS data during data analysis. A review of hybrid EEG-fNIRS in the context of BCI including preprocessing steps and data analysis can be found in (Ahn and Jun 2017; Li et al. 2022; Z. Liu et al. 2021).

Chapter 5

Appendix

5.1 CSP

CSP finds a linear transformation of two covariance matrices that maximizes the ratio of the variances between two classes of signals for discrimination. Consider here the case of binary classification. Roughly, CSP can be seen as the supervised version of principal component analysis (PCA) by changing from the singular value decomposition for the overall covariance matrix to the generalized eigenvalue problem (GEP) for a pair of covariance matrices (A,B), one for each class (Blankertz et al. 2008). The following derivations are from (Ghojogh, Karray, and Crowley 2019; He and D. Wu 2018).

Let $X_{i,j} \in \mathbb{R}^{C \times T}$ be a segment of the ith EEG trial belonging to class j where $j \in 1,2,3$, where C is the number of channels and T is the number of sampled time-points. X_i is assumed to be already zero-centered and scaled. The mean covariance matrix for each class can then be estimated as follow:

$$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^{N_j} X_{i,j} X_{i.j}^T \tag{5.1}$$

where N_j is the total number of trial belonging to class j. Assume here that Σ_1 and Σ_2 are invertible. For binary classification $j = 1,2$, we seek to identify a set of spatial filters W that optimize the following objective function:

$$\underset{W=[w_1,\ldots,w'_C]}{\arg\max} = \sum_{i=1}^{\frac{C'}{2}} \frac{w_i^T \Sigma_1 w_i}{w_i^T \Sigma_2 w_i} + \sum_{i=\frac{C'}{2}+1}^{C'} \frac{w_i^T \Sigma_2 w_i}{w_i^T \Sigma_1 w_i} \tag{5.2}$$

which is equivalent to picking the top C'/2 spatial filters for the following objective function:

$$\underset{W=[w_1,\ldots,w'_C]}{\arg\max} = \frac{w_i^T \Sigma_1 w_i}{w_i^T \Sigma_2 w_i} \tag{5.3}$$

and similarly, the top C'/2 spatial filters for the following:

$$\underset{W=[w_1,\ldots,w'_C]}{\arg\max} = \frac{w_i^T \Sigma_2 w_i}{w_i^T \Sigma_1 w_i} \tag{5.4}$$

The Lagrangian function for 5.3 is:

$$\mathcal{L}(w,\lambda) = w^T \Sigma_1 w - \lambda(w^T \Sigma_2 w - 1) \tag{5.5}$$

Setting the derivative of $\mathcal{L}$ with respect to w to 0, we obtain the following:

$$\frac{\partial \mathcal{L}}{\partial w} = 2\Sigma_1 w - 2\lambda\Sigma_2 w \overset{set}{=} 0 \tag{5.6}$$

$$\Rightarrow 2\Sigma_1 w = 2\lambda\Sigma_2 w \Rightarrow \Sigma_1 w = \lambda\Sigma_2 w \tag{5.7}$$

which we recognize as the GEP. If Σ_2 is invertible, then it can be simplified into standard eigenvalue problem for the matrix $\Sigma_2^{-1}\Sigma_1$. By assumption that both Σ_1 and Σ_2 are invertible, $\Sigma_2^{-1}\Sigma_1$ and $\Sigma_1^{-1}\Sigma_2$ are just inverse of each other and hence will yield identical set of eigenvectors. Thus equation 5.3 and 5.4 can be simplified to equation 5.3 where after solving the optimization problem, we ordered the eigenvalues λ_i and their corresponding eigenvectors $w_i, i = 1,\ldots C$. We next pick the top $\frac{C'}{2}$ eigenvectors and bottom $\frac{C'}{2}$ eigenvectors to construct $W \in \mathbb{R}^{C'\times C}(C' < C)$. In turn, W acts as a dimension reduction from C-dimension to C'-dimension:

$$X_{i,j}^{CSP} = W^T X_{i,j} \in \mathbb{R}^{C'\times T} \tag{5.8}$$

Lastly, since $X_{i,j}$ is already whitened (centered and scaled), the GEP can be solved in MATLAB as ($W = eig(\Sigma_1, \Sigma_1 + \Sigma_2)$). Even when Σ_1 and Σ_2 are not invertible, W can still be obtained from equation 5.3 by using the GEP formulation but the ordering of the GEP

pair matters.

Multiclass CSP. For the 3-class classification, the problem become that of simultaneously diagonalizing multiple standard eigenvalue problems:

$$\begin{aligned} W^T\Sigma_1 W &= \Lambda_1 \\ W^T\Sigma_2 W &= \Lambda_2 \\ W^T\Sigma_3 W &= \Lambda_3 \end{aligned} \tag{5.9}$$

No close-form analytical solution exists for the simultaneous diagonalization of more than 2 square matrices and thus we resorted to joint approximate diagonalization (JAD), a class of numerical algorithms which aim to identify W that makes Λ_1,Λ_2 and Λ_3 as diagonal as possible. More precisely, (Dinh-Tuan Pham and Cardoso 2001) introduced the diagonalization metric as the minimization of the following nonconvex objective function:

$$\mathcal{L}(W) = \frac{1}{2n}\sum_{i=1}^{3}[\log\det \operatorname{diag}(W^T\Sigma_i W) - \log\det(W^T\Sigma_i W)] \tag{5.10}$$

where W is the diagonalizing matrix. (Ablin, Cardoso, and Gramfort 2018) proposed the use of 2^{nd}-order optimization, termed quasi-Newton's method. This requires calculating the Hessian matrix which is computationally intensive. They instead estimate the sparse Hessian matrix before performing the line search. The line search converges quadratically. Interested readers can refer to aforementioned reference for more details. In addition, the codes are provided by the authors.

Multi-class CSP. In addition, we cannot use the objective function in 5.2 for multiclass classification, which provide a way to select C' out of C spatial vectors from W. Instead, this study employs information-theoretic based algorithm, which was proposed by (Grosse-Wentrup and Buss 2008) in his 2008 paper. The algorithm, along with its assumptions and some notes will be briefly discussed here. We seek to identify w that maximize the mutual

information between the class distribution j and the distribution of a filtered data:

$$\arg\max_{w} I(j, w^T X) \tag{5.11}$$

where X is the random variable of the EEG data. The distribution of X given class label j is assumed to be Gaussian with zero mean and a covariance matrix Σ_j. Even though we are computing the entropy between a discrete and a continuous random variable, we can use differential entropy for X which is needed for expanding the mutual information as follows:

$$\begin{aligned} I(j, w^T X) &= H(w^T X) - H(x^T X | j) \\ &= H(w^T X) - \sum_{i=1}^{3} P(j_i) \log \sqrt{2\pi e w^T \Sigma_i w} \end{aligned} \tag{5.12}$$

where the marginal entropy $H(w^T X)$ is approximate as

$$H(w^T X) \approx \log \sqrt{2\pi e} - \frac{3}{16} \Big(\sum_{i=1}^{3} P(j_i)((w^T \Sigma_i w)^2 - 1) \Big)^2 \tag{5.13}$$

Thus, after neglecting constant, we have

$$I(j, w^T X) \approx - \sum_{i=1}^{3} P(j_i) \log \sqrt{w^T \Sigma_i w} - \frac{3}{16} \Big(\sum_{i=1}^{3} P(j_i)((w^T \Sigma_i w)^2 - 1) \Big)^2 \tag{5.14}$$

Equation 5.14 is used to compute the mutual information for each spatial filter w. We picked C' filters with the highest mutual information and finally use equation 5.8 to get the filtered signals.

Bibliography

[1] Ablin, P., Cardoso, J.-F., and Gramfort, A. "Beyond Pham's algorithm for joint diagonalization". In: (2018). DOI: 10.48550/ARXIV.1811.11433. URL: https://arxiv.org/abs/1811.11433 (visited on 10/04/2022).

[2] Ahn, S. and Jun, S. C. "Multi-Modal Integration of EEG-fNIRS for Brain-Computer Interfaces – Current Limitations and Future Directions". In: *Frontiers in Human Neuroscience* 11 (Oct. 2017), p. 503. ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00503. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00503/full (visited on 11/21/2022).

[3] Ahveninen, J. et al. "Dynamic Oscillatory Processes Governing Cued Orienting and Allocation of Auditory Attention". en. In: *Journal of Cognitive Neuroscience* 25.11 (Nov. 2013), pp. 1926–1943. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/jocn_a_00452. URL: https://direct.mit.edu/jocn/article/25/11/1926/5349/Dynamic-Oscillatory-Processes-Governing-Cued (visited on 10/20/2022).

[4] Ai, G. et al. "Direction and viewing area-sensitive influence of EOG artifacts revealed in the EEG topographic pattern analysis". en. In: *Cognitive Neurodynamics* 10.4 (Aug. 2016), pp. 301–314. ISSN: 1871-4080, 1871-4099. DOI: 10.1007/s11571-016-9382-4. URL: http://link.springer.com/10.1007/s11571-016-9382-4 (visited on 10/26/2022).

[5] Alickovic, E. et al. "A Tutorial on Auditory Attention Identification Methods". In: *Frontiers in Neuroscience* 13 (Mar. 2019), p. 153. ISSN: 1662-453X. DOI: 10.3389/fnins.2019.00153. URL: https://www.frontiersin.org/article/10.3389/fnins.2019.00153/full (visited on 07/07/2022).

[6] Arnal, L. H., Doelling, K. B., and Poeppel, D. "Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy". en. In: *Cerebral Cortex* 25.9 (Sept. 2015), pp. 3077–3085. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhu103. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhu103 (visited on 09/30/2022).

[7] Arnal, L. H. and Giraud, A.-L. "Cortical oscillations and sensory predictions". en. In: *Trends in Cognitive Sciences* 16.7 (July 2012), pp. 390–398. ISSN: 13646613. DOI: 10.1016/j.tics.2012.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661312001210 (visited on 10/14/2022).

[8] Aroudi, A. et al. “Impact of Different Acoustic Components on EEG-Based Auditory Attention Decoding in Noisy and Reverberant Conditions”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.4 (Apr. 2019), pp. 652–663. ISSN: 1534-4320, 1558-0210. DOI: 10.1109/TNSRE.2019.2903404. URL: https://ieeexplore.ieee.org/document/8662636/ (visited on 07/07/2022).

[9] Banerjee, S. et al. “Oscillatory Alpha-Band Mechanisms and the Deployment of Spatial Attention to Anticipated Auditory and Visual Target Locations: Supramodal or Sensory-Specific Control Mechanisms?” en. In: *Journal of Neuroscience* 31.27 (July 2011), pp. 9923–9932. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4660-10.2011. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4660-10.2011 (visited on 10/20/2022).

[10] Barone, P. et al. “Laminar Distribution of Neurons in Extrastriate Areas Projecting to Visual Areas V1 and V4 Correlates with the Hierarchical Rank and Indicates the Operation of a Distance Rule”. en. In: *The Journal of Neuroscience* 20.9 (May 2000), pp. 3263–3281. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.20-09-03263.2000. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.20-09-03263.2000 (visited on 10/20/2022).

[11] Bauer, M., Stenner, M.-P., et al. “Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes”. en. In: *Journal of Neuroscience* 34.48 (Nov. 2014), pp. 16117–16125. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3474-13.2014. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3474-13.2014 (visited on 10/25/2022).

[12] Bauer, M., Kluge, C., et al. “Cholinergic Enhancement of Visual Attention and Neural Oscillations in the Human Brain”. en. In: *Current Biology* 22.5 (Mar. 2012), pp. 397–402. ISSN: 09609822. DOI: 10.1016/j.cub.2012.01.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S0960982212000231 (visited on 10/25/2022).

[13] Bengson, J. J., Mangun, G. R., and Mazaheri, A. “The neural markers of an imminent failure of response inhibition”. en. In: *NeuroImage* 59.2 (Jan. 2012), pp. 1534–1539. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.08.034. URL: https://linkinghub.elsevier.com/retrieve/pii/S105381191100927X (visited on 10/25/2022).

[14] Bertrand, O. et al. “Auditory induced 40-Hz activity during a frequency discrimination task”. en. In: *NeuroImage* 7.4 (May 1998), S370. ISSN: 10538119. DOI: 10.1016/S1053-8119(18)31203-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811918312035 (visited on 09/30/2022).

[15] Biau, E. and Kotz, S. A. “Lower Beta: A Central Coordinator of Temporal Prediction in Multimodal Speech”. In: *Frontiers in Human Neuroscience* 12 (Oct. 2018), p. 434. ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00434. URL: https:

//www.frontiersin.org/article/10.3389/fnhum.2018.00434/full (visited on 09/30/2022).

[16] Bizley, J. K., Nodal, F. R., et al. "Physiological and Anatomical Evidence for Multisensory Interactions in Auditory Cortex". en. In: *Cerebral Cortex* 17.9 (Sept. 2007), pp. 2172–2189. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhl128. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhl128 (visited on 07/07/2022).

[17] Bizley, J. K., Maddox, R. K., and Lee, A. K. "Defining Auditory-Visual Objects: Behavioral Tests and Physiological Mechanisms". en. In: *Trends in Neurosciences* 39.2 (Feb. 2016), pp. 74–85. ISSN: 01662236. DOI: 10.1016/j.tins.2015.12.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0166223615002817 (visited on 07/07/2022).

[18] Blankertz, B. et al. "Optimizing Spatial filters for Robust EEG Single-Trial Analysis". In: *IEEE Signal Processing Magazine* 25.1 (2008), pp. 41–56. ISSN: 1053-5888. DOI: 10.1109/MSP.2008.4408441. URL: http://ieeexplore.ieee.org/document/4408441/ (visited on 10/14/2022).

[19] Boser, B. E., Guyon, I. M., and Vapnik, V. N. "A training algorithm for optimal margin classifiers". en. In: *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*. Pittsburgh, Pennsylvania, United States: ACM Press, 1992, pp. 144–152. ISBN: 978-0-89791-497-0. DOI: 10.1145/130385.130401. URL: http://portal.acm.org/citation.cfm?doid=130385.130401 (visited on 07/08/2022).

[20] Braga, R. M. et al. "Eye Movements during Auditory Attention Predict Individual Differences in Dorsal Attention Network Activity". In: *Frontiers in Human Neuroscience* 10 (May 2016). ISSN: 1662-5161. DOI: 10.3389/fnhum.2016.00164. URL: http://journal.frontiersin.org/Article/10.3389/fnhum.2016.00164/abstract (visited on 10/21/2022).

[21] Bressler, S. L. and Richter, C. G. "Interareal oscillatory synchronization in top-down neocortical processing". en. In: *Current Opinion in Neurobiology* 31 (Apr. 2015), pp. 62–66. ISSN: 09594388. DOI: 10.1016/j.conb.2014.08.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S095943881400172X (visited on 09/30/2022).

[22] Buffalo, E. A. et al. "Laminar differences in gamma and alpha coherence in the ventral stream". en. In: *Proceedings of the National Academy of Sciences* 108.27 (July 2011), pp. 11262–11267. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1011284108. URL: https://pnas.org/doi/full/10.1073/pnas.1011284108 (visited on 10/20/2022).

[23] Busse, L. et al. “The spread of attention across modalities and space in a multisensory object”. en. In: *Proceedings of the National Academy of Sciences* 102.51 (Dec. 2005), pp. 18751–18756. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0507704102. URL: https://pnas.org/doi/full/10.1073/pnas.0507704102 (visited on 07/07/2022).

[24] Calvert, G. A., Brammer, M. J., et al. “Response amplification in sensory-specific cortices during crossmodal binding:” en. In: *NeuroReport* 10.12 (Aug. 1999), pp. 2619–2623. ISSN: 0959-4965. DOI: 10.1097/00001756-199908200-00033. URL: http://journals.lww.com/00001756-199908200-00033 (visited on 07/07/2022).

[25] Calvert, G. A., Hansen, P. C., et al. “Detection of Audio-Visual Integration Sites in Humans by Application of Electrophysiological Criteria to the BOLD Effect”. en. In: *NeuroImage* 14.2 (Aug. 2001), pp. 427–438. ISSN: 10538119. DOI: 10.1006/nimg.2001.0812. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811901908122 (visited on 07/07/2022).

[26] Capotosto, P. et al. “Frontoparietal Cortex Controls Spatial Attention through Modulation of Anticipatory Alpha Rhythms”. en. In: *Journal of Neuroscience* 29.18 (May 2009), pp. 5863–5872. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0539-09.2009. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0539-09.2009 (visited on 10/25/2022).

[27] Carlson, B. “Covariance matrix estimation errors and diagonal loading in adaptive arrays”. In: *IEEE Transactions on Aerospace and Electronic Systems* 24.4 (July 1988), pp. 397–401. ISSN: 00189251. DOI: 10.1109/7.7181. URL: http://ieeexplore.ieee.org/document/7181/ (visited on 09/23/2022).

[28] Chalk, M. et al. “Attention Reduces Stimulus-Driven Gamma Frequency Oscillations and Spike Field Coherence in V1”. en. In: *Neuron* 66.1 (Apr. 2010), pp. 114–125. ISSN: 08966273. DOI: 10.1016/j.neuron.2010.03.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627310001844 (visited on 10/20/2022).

[29] Chen, L.-C., Sandmann, P., et al. “Association of Concurrent fNIRS and EEG Signatures in Response to Auditory and Visual Stimuli”. en. In: *Brain Topography* 28.5 (Sept. 2015), pp. 710–725. ISSN: 0896-0267, 1573-6792. DOI: 10.1007/s10548-015-0424-8. URL: http://link.springer.com/10.1007/s10548-015-0424-8 (visited on 11/21/2022).

[30] Chen, L.-C., Stropahl, M., et al. “Enhanced visual adaptation in cochlear implant users revealed by concurrent EEG-fNIRS”. en. In: *NeuroImage* 146 (Feb. 2017), pp. 600–608. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2016.09.033. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811916304967 (visited on 11/21/2022).

[31] Cheveigné, A. de et al. “Decoding the auditory brain with canonical component analysis”. en. In: *NeuroImage* 172 (May 2018), pp. 206–216. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2018.01.033. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811918300338 (visited on 07/07/2022).

[32] Chi, E. C. and Lange, K. “Stable estimation of a covariance matrix guided by nuclear norm penalties”. en. In: *Computational Statistics & Data Analysis* 80 (Dec. 2014), pp. 117–128. ISSN: 01679473. DOI: 10.1016/j.csda.2014.06.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167947314001893 (visited on 07/08/2022).

[33] Ciccarelli, G. et al. “Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods”. en. In: *Scientific Reports* 9.1 (Dec. 2019), p. 11538. ISSN: 2045-2322. DOI: 10.1038/s41598-019-47795-0. URL: http://www.nature.com/articles/s41598-019-47795-0 (visited on 07/07/2022).

[34] Cómez, C. M. et al. “Frequency Analysis of the Eeg During Spatial Selective Attention”. en. In: *International Journal of Neuroscience* 95.1-2 (Jan. 1998), pp. 17–32. ISSN: 0020-7454, 1543-5245. DOI: 10.3109/00207459809000646. URL: http://www.tandfonline.com/doi/full/10.3109/00207459809000646 (visited on 10/21/2022).

[35] Cortes, C. and Vapnik, V. “Support-vector networks”. en. In: *Machine Learning* 20.3 (Sept. 1995), pp. 273–297. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/BF00994018. URL: http://link.springer.com/10.1007/BF00994018 (visited on 07/08/2022).

[36] Crone, N. E. et al. “Induced electrocorticographic gamma activity during auditory perception”. en. In: *Clinical Neurophysiology* 112.4 (Apr. 2001), pp. 565–582. ISSN: 13882457. DOI: 10.1016/S1388-2457(00)00545-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245700005459 (visited on 09/30/2022).

[37] Deng, Y., Choi, I., and Shinn-Cunningham, B. “Topographic specificity of alpha power during auditory spatial attention”. en. In: *NeuroImage* 207 (Feb. 2020), p. 116360. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2019.116360. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811919309516 (visited on 09/23/2022).

[38] Deouell, L. Y. et al. “Cerebral Responses to Change in Spatial Location of Unattended Sounds”. en. In: *Neuron* 55.6 (Sept. 2007), pp. 985–996. ISSN: 08966273. DOI: 10.1016/j.neuron.2007.08.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627307006484 (visited on 07/07/2022).

[39] Desimone, R. and Duncan, J. “Neural Mechanisms of Selective Visual Attention”. en. In: *Annual Review of Neuroscience* 18.1 (Mar. 1995), pp. 193–222. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev.ne.18.030195.001205. URL: https://www.annualreviews.org/doi/10.1146/annurev.ne.18.030195.001205 (visited on 07/07/2022).

[40] Dingcheng Feng, Feng Chen, and Wenli Xu. “Efficient leave-one-out strategy for supervised feature selection”. In: *Tsinghua Science and Technology* 18.6 (Dec. 2013), pp. 629–635. ISSN: 1007-0214. DOI: 10.1109/TST.2013.6678908. URL: http://ieeexplore.ieee.org/document/6678908/ (visited on 07/08/2022).

[41] Dinh-Tuan Pham and Cardoso, J.-F. “Blind separation of instantaneous mixtures of nonstationary sources”. In: *IEEE Transactions on Signal Processing* 49.9 (Sept. 2001), pp. 1837–1848. ISSN: 1053587X. DOI: 10.1109/78.942614. URL: http://ieeexplore.ieee.org/document/942614/ (visited on 10/04/2022).

[42] Donner, T. H. and Siegel, M. “A framework for local cortical oscillation patterns”. en. In: *Trends in Cognitive Sciences* 15.5 (May 2011), pp. 191–199. ISSN: 13646613. DOI: 10.1016/j.tics.2011.03.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661311000428 (visited on 10/14/2022).

[43] Drew, P. J. “Vascular and neural basis of the BOLD signal”. en. In: *Current Opinion in Neurobiology* 58 (Oct. 2019), pp. 61–69. ISSN: 09594388. DOI: 10.1016/j.conb.2019.06.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959438818302575 (visited on 10/05/2022).

[44] Farah, M. J. et al. “Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal?” en. In: *Neuropsychologia* 27.4 (Jan. 1989), pp. 461–470. ISSN: 00283932. DOI: 10.1016/0028-3932(89)90051-1. URL: https://linkinghub.elsevier.com/retrieve/pii/0028393289900511 (visited on 10/20/2022).

[45] Fell, J., Hinrichs, H., and Röschke, J. “Time course of human 40 Hz EEG activity accompanying P3 responses in an auditory oddball paradigm”. en. In: *Neuroscience Letters* 235.3 (Oct. 1997), pp. 121–124. ISSN: 03043940. DOI: 10.1016/S0304-3940(97)00730-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304394097007301 (visited on 09/30/2022).

[46] Fix, E. and Hodges, J. L. “Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties”. In: *International Statistical Review / Revue Internationale de Statistique* 57.3 (Dec. 1989), p. 238. ISSN: 03067734. DOI: 10.2307/1403797. URL: https://www.jstor.org/stable/1403797?origin=crossref (visited on 07/08/2022).

[47] Frey, J. N. et al. “Selective Modulation of Auditory Cortical Alpha Activity in an Audiovisual Spatial Attention Task”. en. In: *Journal of Neuroscience* 34.19 (May 2014), pp. 6634–6639. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4813-13.2014. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4813-13.2014 (visited on 10/14/2022).

[48] Friedman, J., Hastie, T., and Tibshirani, R. “Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)”. In: *The Annals of Statistics* 28.2 (Apr. 2000). ISSN: 0090-5364. DOI: 10.1214/aos/1016218223. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-28/issue-2/Additive-logistic-regression--a-statistical-view-of-boosting-With/10.1214/aos/1016218223.full (visited on 07/08/2022).

[49] Fries, P. “Rhythms for Cognition: Communication through Coherence”. en. In: *Neuron* 88.1 (Oct. 2015), pp. 220–235. ISSN: 08966273. DOI: 10.1016/j.neuron.2015.09.034. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627315008235 (visited on 10/20/2022).

[50] Fujioka, T. et al. “Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations”. en. In: *Journal of Neuroscience* 32.5 (Feb. 2012), pp. 1791–1802. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4107-11.2012. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4107-11.2012 (visited on 10/14/2022).

[51] Fujioka, T. et al. “Beta and Gamma Rhythms in Human Auditory Cortex during Musical Beat Processing”. en. In: *Annals of the New York Academy of Sciences* 1169.1 (July 2009), pp. 89–92. ISSN: 00778923. DOI: 10.1111/j.1749-6632.2009.04779.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1749-6632.2009.04779.x (visited on 10/14/2022).

[52] Gagnon, L. et al. “Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling”. en. In: *NeuroImage* 56.3 (June 2011), pp. 1362–1371. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.03.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811911002606 (visited on 08/08/2022).

[53] Gao, Y. et al. “Selective Attention Enhances Beta-Band Cortical Oscillation to Speech under “Cocktail-Party” Listening Conditions”. In: *Frontiers in Human Neuroscience* 11 (Feb. 2017). ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00034. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2017.00034/full (visited on 09/23/2022).

[54] Garg, A., Schwartz, D., and Stevens, A. A. “Orienting auditory spatial attention engages frontal eye fields and medial occipital cortex in congenitally blind humans”. en. In: *Neuropsychologia* 45.10 (2007), pp. 2307–2321. ISSN: 00283932. DOI: 10.

1016/j.neuropsychologia.2007.02.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0028393207000735 (visited on 11/17/2022).

[55] Gasser, T., Sroka, L., and Möcks, J. "The transfer of EOG activity into the EEG for eyes open and closed". en. In: *Electroencephalography and Clinical Neurophysiology* 61.2 (Aug. 1985), pp. 181–193. ISSN: 00134694. DOI: 10.1016/0013-4694(85)91058-2. URL: https://linkinghub.elsevier.com/retrieve/pii/0013469485910582 (visited on 09/23/2022).

[56] Geirnaert, S., Francart, T., and Bertrand, A. "Fast EEG-Based Decoding Of The Directional Focus Of Auditory Attention Using Common Spatial Patterns". In: *IEEE Transactions on Biomedical Engineering* 68.5 (May 2021), pp. 1557–1568. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2020.3033446. URL: https://ieeexplore.ieee.org/document/9238421/ (visited on 07/07/2022).

[57] Geirnaert, S., Vandecappelle, S., et al. "Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices". In: *IEEE Signal Processing Magazine* 38.4 (July 2021), pp. 89–102. ISSN: 1053-5888, 1558-0792. DOI: 10.1109/MSP.2021.3075932. URL: https://ieeexplore.ieee.org/document/9467380/ (visited on 07/07/2022).

[58] Ghojogh, B., Karray, F., and Crowley, M. "Eigenvalue and Generalized Eigenvalue Problems: Tutorial". In: (2019). DOI: 10.48550/ARXIV.1903.11240. URL: https://arxiv.org/abs/1903.11240 (visited on 11/17/2022).

[59] Giannitrapani, D. "Scanning mechanisms and the EEG". en. In: *Electroencephalography and Clinical Neurophysiology* 30.2 (Feb. 1971), pp. 139–146. ISSN: 00134694. DOI: 10.1016/0013-4694(71)90274-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0013469471902744 (visited on 10/21/2022).

[60] Goldman-Rakic, P. S. and Rakic, P. "Preface: Cerebral Cortex Has Come of Age". en. In: *Cerebral Cortex* 1.1 (Jan. 1991), pp. 1–1. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/1.1.1-a. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/1.1.1-a (visited on 10/20/2022).

[61] Green, D. M. and Swets, J. A. *Signal detection theory and psychophysics*. eng. Repr. ed. Los Altos Hills, Calif: Peninsula Publ, 2000. ISBN: 978-0-932146-23-6.

[62] Grent-'t-Jong, T. et al. "Differential Functional Roles of Slow-Wave and Oscillatory-Alpha Activity in Visual Sensory Cortex during Anticipatory Visual–Spatial Attention". en. In: *Cerebral Cortex* 21.10 (Oct. 2011), pp. 2204–2216. ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bhq279. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhq279 (visited on 10/25/2022).

[63] Gross, J. et al. “Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain”. en. In: *PLoS Biology* 11.12 (Dec. 2013). Ed. by D. Poeppel, e1001752. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001752. URL: https://dx.plos.org/10.1371/journal.pbio.1001752 (visited on 10/24/2022).

[64] Grosse-Wentrup, M. and Buss, M. “Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction”. In: *IEEE Transactions on Biomedical Engineering* 55.8 (Aug. 2008), pp. 1991–2000. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2008.921154. URL: https://ieeexplore.ieee.org/document/4473042/ (visited on 10/14/2022).

[65] Gruber, T. “Selective visual-spatial attention alters induced gamma band responses in the human EEG”. In: *Clinical Neurophysiology* 110.12 (Dec. 1999), pp. 2074–2085. ISSN: 13882457. DOI: 10.1016/S1388-2457(99)00176-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245799001765 (visited on 10/27/2022).

[66] Gurtubay, I. et al. “Gamma band activity in an auditory oddball paradigm studied with the wavelet transform”. en. In: *Clinical Neurophysiology* 112.7 (July 2001), pp. 1219–1228. ISSN: 13882457. DOI: 10.1016/S1388-2457(01)00557-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245701005570 (visited on 09/30/2022).

[67] Hagemann, D. and Naumann, E. “The effects of ocular artifacts on (lateralized) broadband power in the EEG”. en. In: *Clinical Neurophysiology* 112.2 (Feb. 2001), pp. 215–231. ISSN: 13882457. DOI: 10.1016/S1388-2457(00)00541-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245700005411 (visited on 09/23/2022).

[68] Haig, A. R., De Pascalis, V., and Gordon, E. “Peak gamma latency correlated with reaction time in a conventional oddball paradigm”. en. In: *Clinical Neurophysiology* 110.1 (Jan. 1999), pp. 158–165. ISSN: 13882457. DOI: 10.1016/S0013-4694(98)00112-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S0013469498001126 (visited on 09/30/2022).

[69] Hämäläinen, M. et al. “Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain”. en. In: *Reviews of Modern Physics* 65.2 (Apr. 1993), pp. 413–497. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.65.413. URL: https://link.aps.org/doi/10.1103/RevModPhys.65.413 (visited on 10/13/2022).

[70] Hanslmayr, S. et al. “Prestimulus oscillations predict visual perception performance between and within subjects”. en. In: *NeuroImage* 37.4 (Oct. 2007), pp. 1465–1473. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2007.07.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811907006039 (visited on 10/14/2022).

[71] He, H. and Wu, D. "Spatial Filtering for Brain Computer Interfaces: A Comparison between the Common Spatial Pattern and Its Variant". In: *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. Qingdao: IEEE, Sept. 2018, pp. 1–6. ISBN: 978-1-5386-7946-3. DOI: 10.1109/ICSPCC.2018.8567789. URL: https://ieeexplore.ieee.org/document/8567789/ (visited on 11/17/2022).

[72] Heeger, D. J. and Ress, D. "What does fMRI tell us about neuronal activity?" en. In: *Nature Reviews Neuroscience* 3.2 (Feb. 2002), pp. 142–151. ISSN: 1471-003X, 1471-0048. DOI: 10.1038/nrn730. URL: http://www.nature.com/articles/nrn730 (visited on 10/04/2022).

[73] Hong, K.-S. and Santosa, H. "Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy". en. In: *Hearing Research* 333 (Mar. 2016), pp. 157–166. ISSN: 03785955. DOI: 10.1016/j.heares.2016.01.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378595515300125 (visited on 07/07/2022).

[74] Jäncke, L., Mirzazade, S., and Joni Shah, N. "Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects". en. In: *Neuroscience Letters* 266.2 (May 1999), pp. 125–128. ISSN: 03043940. DOI: 10.1016/S0304-3940(99)00288-8. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304394099002888 (visited on 07/08/2022).

[75] Jerbi, K. et al. "Task-related gamma-band dynamics from an intracerebral perspective: Review and implications for surface EEG and MEG". en. In: *Human Brain Mapping* 30.6 (June 2009), pp. 1758–1771. ISSN: 10659471, 10970193. DOI: 10.1002/hbm.20750. URL: https://onlinelibrary.wiley.com/doi/10.1002/hbm.20750 (visited on 09/30/2022).

[76] Jia, X., Tanabe, S., and Kohn, A. "Gamma and the Coordination of Spiking Activity in Early Visual Cortex". en. In: *Neuron* 77.4 (Feb. 2013), pp. 762–774. ISSN: 08966273. DOI: 10.1016/j.neuron.2012.12.036. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627313000445 (visited on 10/20/2022).

[77] Joliot, M., Ribary, U., and Llinás, R. "Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding." en. In: *Proceedings of the National Academy of Sciences* 91.24 (Nov. 1994), pp. 11748–11751. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.91.24.11748. URL: https://pnas.org/doi/full/10.1073/pnas.91.24.11748 (visited on 09/30/2022).

[78] Kaiser, J. and Lutzenberger, W. "Human gamma-band activity: A window to cognitive processing". In: *NeuroReport* 16.3 (Feb. 2005), pp. 207–211. DOI: 10.1097/00001756-200502280-00001. URL: https://journals.lww.com/neuroreport/Fulltext/2005/02280/Combined_EEG_

and _ MEG _ recordings _ of _ visual _ 40 _ Hz . 1 . aspx ? casa _ token = U6FW7pnXsOUAAAAA : JImHGhOaco - hQbhfBzrO9Cee7MwG9oHFWdblrD _ 9S4emfUQNOCYNLxmo3tgIjKrVLnFMnQE-4Joa2S85pYjhzLi9I2k.

[79] Kaiser, J. and Lutzenberger, W. "Induced Gamma-Band Activity and Human Brain Function". en. In: *The Neuroscientist* 9.6 (Dec. 2003), pp. 475–484. ISSN: 1073-8584, 1089-4098. DOI: 10.1177/1073858403259137. URL: http://journals.sagepub.com/doi/10.1177/1073858403259137 (visited on 10/03/2022).

[80] Kamiński, J. et al. "Beta band oscillations engagement in human alertness process". en. In: *International Journal of Psychophysiology* 85.1 (July 2012), pp. 125–128. ISSN: 01678760. DOI: 10.1016/j.ijpsycho.2011.11.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167876011003539 (visited on 10/21/2022).

[81] Katsuki, F. and Constantinidis, C. "Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems". en. In: *The Neuroscientist* 20.5 (Oct. 2014), pp. 509–521. ISSN: 1073-8584, 1089-4098. DOI: 10.1177/1073858413514136. URL: http://journals.sagepub.com/doi/10.1177/1073858413514136 (visited on 11/17/2022).

[82] Kayser, C., Petkov, C. I., and Logothetis, N. K. "Visual Modulation of Neurons in Auditory Cortex". en. In: *Cerebral Cortex* 18.7 (July 2008), pp. 1560–1574. ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bhm187. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhm187 (visited on 07/07/2022).

[83] Kelly, S. P., Gomez-Ramirez, M., and Foxe, J. J. "The strength of anticipatory spatial biasing predicts target discrimination at attended locations: a high-density EEG study". en. In: *European Journal of Neuroscience* 30.11 (Dec. 2009), pp. 2224–2234. ISSN: 0953816X, 14609568. DOI: 10.1111/j.1460-9568.2009.06980.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2009.06980.x (visited on 10/25/2022).

[84] Kelly, S. P., Lalor, E. C., et al. "Increases in Alpha Oscillatory Power Reflect an Active Retinotopic Mechanism for Distracter Suppression During Sustained Visuospatial Attention". en. In: *Journal of Neurophysiology* 95.6 (June 2006), pp. 3844–3851. ISSN: 0022-3077, 1522-1598. DOI: 10.1152/jn.01234.2005. URL: https://www.physiology.org/doi/10.1152/jn.01234.2005 (visited on 10/14/2022).

[85] Kerkoerle, T. van et al. "Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex". en. In: *Proceedings of the National Academy of Sciences* 111.40 (Oct. 2014), pp. 14332–14341. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1402773111. URL: https://pnas.org/doi/full/10.1073/pnas.1402773111 (visited on 10/20/2022).

[86] Kerlin, J. R., Shahin, A. J., and Miller, L. M. “Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party"”. en. In: *Journal of Neuroscience* 30.2 (Jan. 2010), pp. 620–628. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3631-09.2010. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3631-09.2010 (visited on 10/20/2022).

[87] Kisley, M. A. and Cornwell, Z. M. “Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation”. en. In: *Clinical Neurophysiology* 117.11 (Nov. 2006), pp. 2549–2563. ISSN: 13882457. DOI: 10.1016/j.clinph.2006.08.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245706014040 (visited on 10/03/2022).

[88] Klatt, L.-I. et al. “The contribution of selective spatial attention to sound detection and sound localization: Evidence from event-related potentials and lateralized alpha oscillations”. en. In: *Biological Psychology* 138 (Oct. 2018), pp. 133–145. ISSN: 03010511. DOI: 10.1016/j.biopsycho.2018.08.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0301051118301352 (visited on 10/20/2022).

[89] Knudsen, E. I. “Fundamental Components of Attention”. en. In: *Annual Review of Neuroscience* 30.1 (July 2007), pp. 57–78. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev.neuro.30.051606.094256. URL: https://www.annualreviews.org/doi/10.1146/annurev.neuro.30.051606.094256 (visited on 07/07/2022).

[90] Lakatos, P. et al. “Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection”. en. In: *Science* 320.5872 (Apr. 2008), pp. 110–113. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1154735. URL: https://www.science.org/doi/10.1126/science.1154735 (visited on 10/14/2022).

[91] Larson, E. and Lee, A. K. “The cortical dynamics underlying effective switching of auditory spatial attention”. en. In: *NeuroImage* 64 (Jan. 2013), pp. 365–370. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.09.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811912009007 (visited on 07/07/2022).

[92] Laurienti, P. J. et al. “Deactivation of Sensory-Specific Cortex by Cross-Modal Stimuli”. en. In: *Journal of Cognitive Neuroscience* 14.3 (Apr. 2002), pp. 420–429. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/089892902317361930. URL: https://direct.mit.edu/jocn/article/14/3/420/3621/Deactivation-of-Sensory-Specific-Cortex-by-Cross (visited on 07/07/2022).

[93] Ledoit, O. and Wolf, M. “A well-conditioned estimator for large-dimensional covariance matrices”. en. In: *Journal of Multivariate Analysis* 88.2 (Feb. 2004), pp. 365–411. ISSN: 0047259X. DOI: 10.1016/S0047-259X(03)00096-4. URL:

https://linkinghub.elsevier.com/retrieve/pii/S0047259X03000964 (visited on 07/08/2022).

[94] Lenz, D. et al. "Human EEG very high frequency oscillations reflect the number of matches with a template in auditory short-term memory". en. In: *Brain Research* 1220 (July 2008), pp. 81–92. ISSN: 00068993. DOI: 10.1016/j.brainres.2007.10.053. URL: https://linkinghub.elsevier.com/retrieve/pii/S000689930702481X (visited on 09/30/2022).

[95] Li, R. et al. "Concurrent fNIRS and EEG for Brain Function Investigation: A Systematic, Methodology-Focused Review". en. In: *Sensors* 22.15 (Aug. 2022), p. 5865. ISSN: 1424-8220. DOI: 10.3390/s22155865. URL: https://www.mdpi.com/1424-8220/22/15/5865 (visited on 11/21/2022).

[96] Lindquist, M. A. et al. "Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling". en. In: *NeuroImage* 45.1 (Mar. 2009), S187–S198. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.10.065. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811908012056 (visited on 07/08/2022).

[97] Liu, H. et al. "Determination of optical properties and blood oxygenation in tissue using continuous NIR light". In: *Physics in Medicine and Biology* 40.11 (Nov. 1995), pp. 1983–1993. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/40/11/015. URL: https://iopscience.iop.org/article/10.1088/0031-9155/40/11/015 (visited on 10/05/2022).

[98] Liu, Y. et al. "Top-down Modulation of Neural Activity in Anticipatory Visual Attention: Control Mechanisms Revealed by Simultaneous EEG-fMRI". en. In: *Cerebral Cortex* (Sept. 2014), bhu204. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhu204. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhu204 (visited on 10/25/2022).

[99] Liu, Z. et al. "A systematic review on hybrid EEG/fNIRS in brain-computer interface". en. In: *Biomedical Signal Processing and Control* 68 (July 2021), p. 102595. ISSN: 17468094. DOI: 10.1016/j.bspc.2021.102595. URL: https://linkinghub.elsevier.com/retrieve/pii/S1746809421001920 (visited on 11/21/2022).

[100] Logothetis, N. K. "The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal". en. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357.1424 (Aug. 2002). Ed. by A. Parker, A. Derrington, and C. Blakemore, pp. 1003–1037. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2002.1114. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2002.1114 (visited on 10/04/2022).

[101] Logothetis, N. K. et al. “Neurophysiological investigation of the basis of the fMRI signal”. en. In: *Nature* 412.6843 (July 2001), pp. 150–157. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/35084005. URL: https://www.nature.com/articles/35084005 (visited on 10/04/2022).

[102] Lühmann, A. von et al. “Using the General Linear Model to Improve Performance in fNIRS Single Trial Analysis and Classification: A Perspective”. In: *Frontiers in Human Neuroscience* 14 (Feb. 2020), p. 30. ISSN: 1662-5161. DOI: 10.3389/fnhum.2020.00030. URL: https://www.frontiersin.org/article/10.3389/fnhum.2020.00030/full (visited on 09/26/2022).

[103] Luke, R. et al. “Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm”. In: *Neurophotonics* 8.02 (May 2021). ISSN: 2329-423X. DOI: 10.1117/1.NPh.8.2.025008. URL: https://www.spiedigitallibrary.org/journals/neurophotonics/volume-8/issue-02/025008/Analysis-methods-for-measuring-passive-auditory-fNIRS-responses-generated-by/10.1117/1.NPh.8.2.025008.full (visited on 07/07/2022).

[104] Maddox, R. K. et al. “Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners”. en. In: *eLife* 4 (Feb. 2015), e04995. ISSN: 2050-084X. DOI: 10.7554/eLife.04995. URL: https://elifesciences.org/articles/04995 (visited on 07/07/2022).

[105] Magazzini, L. and Singh, K. D. “Spatial attention modulates visual gamma oscillations across the human ventral stream”. en. In: *NeuroImage* 166 (Feb. 2018), pp. 219–229. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2017.10.069. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811917309072 (visited on 10/25/2022).

[106] Markov, N. T. et al. “Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex”. en. In: *Journal of Comparative Neurology* 522.1 (Jan. 2014), pp. 225–259. ISSN: 0021-9967, 1096-9861. DOI: 10.1002/cne.23458. URL: https://onlinelibrary.wiley.com/doi/10.1002/cne.23458 (visited on 10/20/2022).

[107] Marshall, L., Mölle, M., and Bartsch, P. “Event-related gamma band activity during passive and active oddball tasks”. In: *NeuroReport* 7.9 (June 1996), pp. 1517–1520.

[108] Marshall, T. R., O’Shea, J., et al. “Frontal Eye Fields Control Attentional Modulation of Alpha and Gamma Oscillations in Contralateral Occipitoparietal Cortex”. en. In: *Journal of Neuroscience* 35.4 (Jan. 2015), pp. 1638–1647. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3116-14.2015. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3116-14.2015 (visited on 10/25/2022).

[109] Marshall, T. R., Bergmann, T. O., and Jensen, O. “Frontoparietal Structural Connectivity Mediates the Top-Down Control of Neuronal Synchronization Associated with Selective Attention”. en. In: *PLOS Biology* 13.10 (Oct. 2015). Ed. by T. Behrens, e1002272. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002272. URL: https://dx.plos.org/10.1371/journal.pbio.1002272 (visited on 10/25/2022).

[110] Martuzzi, R. et al. “Multisensory Interactions within Human Primary Cortices Revealed by BOLD Dynamics”. en. In: *Cerebral Cortex* 17.7 (July 2007), pp. 1672–1679. ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bhl077. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhl077 (visited on 07/07/2022).

[111] Michalareas, G. et al. “Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas”. en. In: *Neuron* 89.2 (Jan. 2016), pp. 384–397. ISSN: 08966273. DOI: 10.1016/j.neuron.2015.12.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627315011204 (visited on 10/20/2022).

[112] Michalka, S. W., Kong, L., et al. “Short-Term Memory for Space and Time Flexibly Recruit Complementary Sensory-Biased Frontal Lobe Attention Networks”. en. In: *Neuron* 87.4 (Aug. 2015), pp. 882–892. ISSN: 08966273. DOI: 10.1016/j.neuron.2015.07.028. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627315006704 (visited on 07/07/2022).

[113] Michalka, S. W., Rosen, M. L., et al. “Auditory Spatial Coding Flexibly Recruits Anterior, but Not Posterior, Visuotopic Parietal Cortex”. en. In: *Cerebral Cortex* 26.3 (Mar. 2016), pp. 1302–1308. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhv303. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhv303 (visited on 07/08/2022).

[114] Mirkovic, B. et al. “Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications”. In: *Journal of Neural Engineering* 12.4 (Aug. 2015), p. 046007. ISSN: 1741-2560, 1741-2552. DOI: 10.1088/1741-2560/12/4/046007. URL: https://iopscience.iop.org/article/10.1088/1741-2560/12/4/046007 (visited on 07/07/2022).

[115] Mock, J. R. et al. “Rapid cortical dynamics associated with auditory spatial attention gradients”. In: *Frontiers in Neuroscience* 9 (June 2015). ISSN: 1662-453X. DOI: 10.3389/fnins.2015.00179. URL: http://journal.frontiersin.org/Article/10.3389/fnins.2015.00179/abstract (visited on 10/27/2022).

[116] Muller, M. et al. “Visually induced gamma-band responses in human electroencephalographic activity ? a link to animal studies”. en. In: *Experimental Brain Research* 112.1 (Nov. 1996). ISSN: 0014-4819, 1432-1106. DOI: 10.1007/BF00227182. URL: http://link.springer.com/10.1007/BF00227182 (visited on 10/24/2022).

[117] Müller, M. M. and Keil, A. "Neuronal Synchronization and Selective Color Processing in the Human Brain". en. In: *Journal of Cognitive Neuroscience* 16.3 (Apr. 2004), pp. 503–522. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/089892904322926827. URL: https://direct.mit.edu/jocn/article/16/3/503/3836/Neuronal-Synchronization-and-Selective-Color (visited on 10/24/2022).

[118] Müller, N. and Weisz, N. "Lateralized Auditory Cortical Alpha Band Activity and Interregional Connectivity Pattern Reflect Anticipation of Target Sounds". en. In: *Cerebral Cortex* 22.7 (July 2012), pp. 1604–1613. ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bhr232. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr232 (visited on 10/14/2022).

[119] Muthukumaraswamy, S. D. "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations". In: *Frontiers in Human Neuroscience* 7 (2013). ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00138. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2013.00138/abstract (visited on 09/29/2022).

[120] Narayanan, A. M. and Bertrand, A. "Analysis of Miniaturization Effects and Channel Selection Strategies for EEG Sensor Networks With Application to Auditory Attention Detection". In: *IEEE Transactions on Biomedical Engineering* 67.1 (Jan. 2020), pp. 234–244. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2019.2911728. URL: https://ieeexplore.ieee.org/document/8693547/ (visited on 07/07/2022).

[121] Ning, M. et al. *Decoding Attended Spatial Location during Complex Scene Analysis with fNIRS*. en. preprint. Bioengineering, Sept. 2022. DOI: 10.1101/2022.09.06.506821. URL: http://biorxiv.org/lookup/doi/10.1101/2022.09.06.506821 (visited on 10/14/2022).

[122] O'Sullivan, A. E., Lim, C. Y., and Lalor, E. C. "Look at me when I'm talking to you: Selective attention at a multisensory cocktail party can be decoded using stimulus reconstruction and alpha power modulations". en. In: *European Journal of Neuroscience* 50.8 (Oct. 2019), pp. 3282–3295. ISSN: 0953-816X, 1460-9568. DOI: 10.1111/ejn.14425. URL: https://onlinelibrary.wiley.com/doi/10.1111/ejn.14425 (visited on 11/21/2022).

[123] O'Sullivan, J. A. et al. "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG". en. In: *Cerebral Cortex* 25.7 (July 2015), pp. 1697–1706. ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bht355. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bht355 (visited on 07/07/2022).

[124] Park, H. et al. “Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners”. en. In: *Current Biology* 25.12 (June 2015), pp. 1649–1653. ISSN: 09609822. DOI: 10.1016/j.cub.2015.04.049. URL: https://linkinghub.elsevier.com/retrieve/pii/S096098221500500X (visited on 10/14/2022).

[125] Parra, L. C. et al. “Recipes for the linear analysis of EEG”. en. In: *NeuroImage* 28.2 (Nov. 2005), pp. 326–341. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2005.05.032. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811905003381 (visited on 10/27/2022).

[126] Petersen, S. E. and Posner, M. I. “The Attention System of the Human Brain: 20 Years After”. en. In: *Annual Review of Neuroscience* 35.1 (July 2012), pp. 73–89. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev-neuro-062111-150525. URL: https://www.annualreviews.org/doi/10.1146/annurev-neuro-062111-150525 (visited on 11/17/2022).

[127] Petkov, C. I. et al. “Attentional modulation of human auditory cortex”. en. In: *Nature Neuroscience* 7.6 (June 2004), pp. 658–663. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/nn1256. URL: http://www.nature.com/articles/nn1256 (visited on 07/08/2022).

[128] Petro, L. S., Paton, A. T., and Muckli, L. “Contextual modulation of primary visual cortex by auditory signals”. en. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1714 (Feb. 2017), p. 20160104. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2016.0104. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0104 (visited on 07/07/2022).

[129] Pollonini, L. et al. “Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy”. en. In: *Hearing Research* 309 (Mar. 2014), pp. 84–93. ISSN: 03785955. DOI: 10.1016/j.heares.2013.11.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378595513002803 (visited on 07/07/2022).

[130] Popov, T. et al. “Brain areas associated with visual spatial attention display topographic organization during auditory spatial attention”. en. In: *Cerebral Cortex* (Aug. 2022), bhac285. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhac285. URL: https://academic.oup.com/cercor/advance-article/doi/10.1093/cercor/bhac285/6668251 (visited on 10/21/2022).

[131] Pulvermüller, F. et al. “High-frequency brain activity: Its possible role in attention, perception and language processing”. en. In: *Progress in Neurobiology* 52.5 (Aug. 1997), pp. 427–445. ISSN: 03010082. DOI: 10.1016/S0301-0082(97)00023-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0301008297000233 (visited on 10/24/2022).

[132] Putze, F. et al. “Hybrid fNIRS-EEG based classification of auditory and visual perception processes”. In: *Frontiers in Neuroscience* 8 (Nov. 2014). ISSN: 1662-453X. DOI: 10.3389/fnins.2014.00373. URL: http://journal.frontiersin.org/article/10.3389/fnins.2014.00373/abstract (visited on 11/21/2022).

[133] Rajagovindan, R. and Ding, M. “From Prestimulus Alpha Oscillation to Visual-evoked Response: An Inverted-U Function and Its Attentional Modulation”. en. In: *Journal of Cognitive Neuroscience* 23.6 (June 2011), pp. 1379–1394. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/jocn.2010.21478. URL: https://direct.mit.edu/jocn/article/23/6/1379/5108/From-Prestimulus-Alpha-Oscillation-to-Visual (visited on 10/25/2022).

[134] Ray, S. et al. “High-frequency gamma activity (80–150Hz) is increased in human cortex during selective attention”. en. In: *Clinical Neurophysiology* 119.1 (Jan. 2008), pp. 116–133. ISSN: 13882457. DOI: 10.1016/j.clinph.2007.09.136. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245707005159 (visited on 10/24/2022).

[135] Richter, C. G. et al. “Top-Down Beta Enhances Bottom-Up Gamma”. en. In: *The Journal of Neuroscience* 37.28 (July 2017), pp. 6698–6711. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3771-16.2017. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3771-16.2017 (visited on 10/20/2022).

[136] Rihs, T., Michel, C., and Thut, G. “A bias for posterior α-band power suppression versus enhancement during shifting versus maintenance of spatial attention”. en. In: *NeuroImage* 44.1 (Jan. 2009), pp. 190–199. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.08.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811908009464 (visited on 10/25/2022).

[137] Rihs, T. A., Michel, C. M., and Thut, G. “Mechanisms of selective inhibition in visual spatial attention are indexed by ?-band EEG synchronization”. en. In: *European Journal of Neuroscience* 25.2 (Jan. 2007), pp. 603–610. ISSN: 0953-816X, 1460-9568. DOI: 10.1111/j.1460-9568.2007.05278.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2007.05278.x (visited on 10/25/2022).

[138] Saarinen, T. et al. “Task-sensitive reconfiguration of corticocortical 6–20 Hz oscillatory coherence in naturalistic human performance”. en. In: *Human Brain Mapping* 36.7 (July 2015), pp. 2455–2469. ISSN: 1065-9471, 1097-0193. DOI: 10.1002/hbm.22784. URL: https://onlinelibrary.wiley.com/doi/10.1002/hbm.22784 (visited on 09/30/2022).

[139] Sakowitz, O. W. et al. “Bisensory stimulation increases gamma-responses over multiple cortical regions”. en. In: *Cognitive Brain Research* 11.2 (Apr. 2001), pp. 267–279. ISSN: 09266410. DOI: 10.1016/S0926-6410(00)00081-1. URL:

https://linkinghub.elsevier.com/retrieve/pii/S0926641000000811 (visited on 10/25/2022).

[140] Sauseng, P., Klimesch, W., et al. "A shift of visual spatial attention is selectively associated with human EEG alpha activity". en. In: *European Journal of Neuroscience* 22.11 (Dec. 2005), pp. 2917–2926. ISSN: 0953816X, 14609568. DOI: 10.1111/j.1460-9568.2005.04482.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2005.04482.x (visited on 10/25/2022).

[141] Sauseng, P., Feldheim, J. F., et al. "Right Prefrontal TMS Disrupts Interregional Anticipatory EEG Alpha Activity during Shifting of Visuospatial Attention". In: *Frontiers in Psychology* 2 (2011). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2011.00241. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00241/abstract (visited on 10/25/2022).

[142] Scholkmann, F. and Wolf, M. "General equation for the differential pathlength factor of the frontal human head depending on wavelength and age". en. In: *Journal of Biomedical Optics* 18.10 (Oct. 2013), p. 105004. ISSN: 1083-3668. DOI: 10.1117/1.JBO.18.10.105004. URL: http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JBO.18.10.105004 (visited on 07/07/2022).

[143] D. L. Schomer and E. Niedermeyer, eds. *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. eng. 6. ed. Philadelphia, Pa. London: Wolters Kluwer, Lippincott Williams & Wilkins, 2011. ISBN: 978-0-7817-8942-4.

[144] Senkowski, D. et al. "Multisensory processing and oscillatory gamma responses: effects of spatial selective attention". en. In: *Experimental Brain Research* 166.3-4 (Oct. 2005), pp. 411–426. ISSN: 0014-4819, 1432-1106. DOI: 10.1007/s00221-005-2381-z. URL: http://link.springer.com/10.1007/s00221-005-2381-z (visited on 10/25/2022).

[145] Serences, J. T. et al. "Coordination of Voluntary and Stimulus-Driven Attentional Control in Human Cortex". en. In: *Psychological Science* 16.2 (Feb. 2005), pp. 114–122. ISSN: 0956-7976, 1467-9280. DOI: 10.1111/j.0956-7976.2005.00791.x. URL: http://journals.sagepub.com/doi/10.1111/j.0956-7976.2005.00791.x (visited on 07/07/2022).

[146] Shomstein, S. "Parietal Cortex Mediates Voluntary Control of Spatial and Nonspatial Auditory Attention". en. In: *Journal of Neuroscience* 26.2 (Jan. 2006), pp. 435–439. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4408-05.2006. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4408-05.2006 (visited on 07/07/2022).

[147] Siegel, M. et al. “Neuronal Synchronization along the Dorsal Visual Pathway Reflects the Focus of Spatial Attention”. en. In: *Neuron* 60.4 (Nov. 2008), pp. 709–719. ISSN: 08966273. DOI: 10.1016/j.neuron.2008.09.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627308007575 (visited on 10/20/2022).

[148] Sinai, A. et al. “Intracranial mapping of auditory perception: Event-related responses and electrocortical stimulation”. en. In: *Clinical Neurophysiology* 120.1 (Jan. 2009), pp. 140–149. ISSN: 13882457. DOI: 10.1016/j.clinph.2008.10.152. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245708012091 (visited on 09/30/2022).

[149] Singh, S. “Magnetoencephalography: Basic principles”. en. In: *Annals of Indian Academy of Neurology* 17.5 (2014), p. 107. ISSN: 0972-2327. DOI: 10.4103/0972-2327.128676. URL: http://www.annalsofian.org/text.asp?2014/17/5/107/128676 (visited on 10/04/2022).

[150] Smith, D. V. et al. “Spatial Attention Evokes Similar Activation Patterns for Visual and Auditory Stimuli”. en. In: *Journal of Cognitive Neuroscience* 22.2 (Feb. 2010), pp. 347–361. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/jocn.2009.21241. URL: https://direct.mit.edu/jocn/article/22/2/347/4797/Spatial-Attention-Evokes-Similar-Activation (visited on 07/28/2022).

[151] Spagna, A., Mackie, M.-A., and Fan, J. “Supramodal executive control of attention”. In: *Frontiers in Psychology* 6 (Feb. 2015). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.00065. URL: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.00065/abstract (visited on 10/21/2022).

[152] Steinmetzger, K. et al. “Auditory cortex activity measured using functional near-infrared spectroscopy (fNIRS) appears to be susceptible to masking by cortical blood stealing”. en. In: *Hearing Research* 396 (Oct. 2020), p. 108069. ISSN: 03785955. DOI: 10.1016/j.heares.2020.108069. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378595520303403 (visited on 07/07/2022).

[153] Strauab, A., Wostmann, M., and Obleser, J. “Cortical alpha oscillations as a tool for auditory selective inhibition”. In: *Frontiers in Human Neuroscience* 8 (May 2014). ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00350. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2014.00350/abstract (visited on 10/14/2022).

[154] Stropahl, M., Chen, L.-C., and Debener, S. “Cortical reorganization in postlingually deaf cochlear implant users: Intra-modal and cross-modal considerations”. en. In: *Hearing Research* 343 (Jan. 2017), pp. 128–137. ISSN: 03785955. DOI: 10.1016/j.heares.2016.07.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378595516301447 (visited on 11/21/2022).

[155] Su, E. et al. “STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG”. In: *IEEE Transactions on Biomedical Engineering* 69.7 (July 2022), pp. 2233–2242. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2022.3140246. URL: https://ieeexplore.ieee.org/document/9669037/ (visited on 07/07/2022).

[156] Sun, L. et al. “Left frontal eye field encodes sound locations during passive listening”. en. In: *Cerebral Cortex* (July 2022), bhac261. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhac261. URL: https://academic.oup.com/cercor/advance-article/doi/10.1093/cercor/bhac261/6647501 (visited on 11/17/2022).

[157] Szabó, B. T., Denham, S. L., and Winkler, I. “Computational Models of Auditory Scene Analysis: A Review”. In: *Frontiers in Neuroscience* 10 (Nov. 2016). ISSN: 1662-453X. DOI: 10.3389/fnins.2016.00524. URL: http://journal.frontiersin.org/article/10.3389/fnins.2016.00524/full (visited on 11/17/2022).

[158] Taillez, T., Kollmeier, B., and Meyer, B. T. “Machine learning for decoding listeners’ attention from electroencephalography evoked by continuous speech”. en. In: *European Journal of Neuroscience* 51.5 (Mar. 2020), pp. 1234–1241. ISSN: 0953-816X, 1460-9568. DOI: 10.1111/ejn.13790. URL: https://onlinelibrary.wiley.com/doi/10.1111/ejn.13790 (visited on 07/07/2022).

[159] Tallon-Baudry, C. “Oscillatory gamma activity in humans and its role in object representation”. In: *Trends in Cognitive Sciences* 3.4 (Apr. 1999), pp. 151–162. ISSN: 13646613. DOI: 10.1016/S1364-6613(99)01299-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661399012991 (visited on 09/30/2022).

[160] Tallon-Baudry, C. et al. “Combined EEG and MEG recordings of visual 40 Hz responses to illusory triangles in human”. In: *NeuroReport* 8.5 (Mar. 1997), pp. 1103–1107. URL: https://journals.lww.com/neuroreport/Fulltext/1997/03240/Combined_EEG_and_MEG_recordings_of_visual_40_Hz.8.aspx.

[161] Thorpe, S., D’Zmura, M., and Srinivasan, R. “Lateralization of Frequency-Specific Networks for Covert Spatial Attention to Auditory Stimuli”. en. In: *Brain Topography* 25.1 (Jan. 2012), pp. 39–54. ISSN: 0896-0267, 1573-6792. DOI: 10.1007/s10548-011-0186-x. URL: http://link.springer.com/10.1007/s10548-011-0186-x (visited on 10/20/2022).

[162] Thut, G. “-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection”. en. In: *Journal of Neuroscience* 26.37 (Sept. 2006), pp. 9494–9502. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0875-06.2006. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0875-06.2006 (visited on 10/25/2022).

[163] Tian, X. et al. “Cerebral Representation of Sound Localization Using Functional Near-Infrared Spectroscopy”. In: *Frontiers in Neuroscience* 15 (Dec. 2021), p. 739706. ISSN: 1662-453X. DOI: 10.3389/fnins.2021.739706. URL: https://www.frontiersin.org/articles/10.3389/fnins.2021.739706/full (visited on 07/07/2022).

[164] Tiitinen, H. T. et al. “Selective attention enhances the auditory 40-Hz transient response in humans”. en. In: *Nature* 364.6432 (July 1993), pp. 59–60. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/364059a0. URL: http://www.nature.com/articles/364059a0 (visited on 09/30/2022).

[165] Tin Kam Ho. “Random decision forests”. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. Montreal, Que., Canada: IEEE Comput. Soc. Press, 1995, pp. 278–282. ISBN: 978-0-8186-7128-9. DOI: 10.1109/ICDAR.1995.598994. URL: http://ieeexplore.ieee.org/document/598994/ (visited on 07/08/2022).

[166] Todorovic, A. et al. “Temporal Expectation and Attention Jointly Modulate Auditory Oscillatory Activity in the Beta Band”. en. In: *PLOS ONE* 10.3 (Mar. 2015). Ed. by C. F. Altmann, e0120288. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0120288. URL: https://dx.plos.org/10.1371/journal.pone.0120288 (visited on 10/14/2022).

[167] Towle, V. L. et al. “ECoG gamma activity during a language task: differentiating expressive and receptive speech areas”. en. In: *Brain* 131.8 (Jan. 2008), pp. 2013–2027. ISSN: 0006-8950, 1460-2156. DOI: 10.1093/brain/awn147. URL: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awn147 (visited on 09/30/2022).

[168] Trapp, G. von et al. “A Decline in Response Variability Improves Neural Signal Detection during Auditory Task Performance”. en. In: *Journal of Neuroscience* 36.43 (Oct. 2016), pp. 11097–11106. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1302-16.2016. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1302-16.2016 (visited on 07/08/2022).

[169] Tune, S., Wöstmann, M., and Obleser, J. “Probing the limits of alpha power lateralisation as a neural marker of selective attention in middle-aged and older listeners”. en. In: *European Journal of Neuroscience* 48.7 (Oct. 2018), pp. 2537–2550. ISSN: 0953816X. DOI: 10.1111/ejn.13862. URL: https://onlinelibrary.wiley.com/doi/10.1111/ejn.13862 (visited on 10/20/2022).

[170] Vandecappelle, S. et al. “EEG-based detection of the locus of auditory attention with convolutional neural networks”. en. In: *eLife* 10 (Apr. 2021), e56481. ISSN: 2050-084X. DOI: 10.7554/eLife.56481. URL: https://elifesciences.org/articles/56481 (visited on 07/07/2022).

[171] Vapnik, V. N. “The Support Vector method”. In: *Artificial Neural Networks — ICANN’97*. Ed. by G. Goos et al. Vol. 1327. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 261–271. ISBN: 978-3-540-63631-1. DOI: 10.1007/BFb0020166. URL: http://link.springer.com/10.1007/BFb0020166 (visited on 07/08/2022).

[172] Vázquez Marrufo, M. et al. “Temporal evolution of α and β bands during visual spatial attention”. en. In: *Cognitive Brain Research* 12.2 (Oct. 2001), pp. 315–320. ISSN: 09266410. DOI: 10.1016/S0926-6410(01)00025-8. URL: https://linkinghub.elsevier.com/retrieve/pii/S0926641001000258 (visited on 10/14/2022).

[173] Watanabe, N. et al. “The relationship between the visually evoked P300 event-related potential and gamma band oscillation in the human medial and basal temporal lobes”. en. In: *Neuroscience Research* 44.4 (Dec. 2002), pp. 421–427. ISSN: 01680102. DOI: 10.1016/S0168-0102(02)00159-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168010202001591 (visited on 09/30/2022).

[174] Weichselgartner, E. and Sperling, G. “Dynamics of Automatic and Controlled Visual Attention”. en. In: *Science* 238.4828 (Nov. 1987), pp. 778–780. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.3672124. URL: https://www.science.org/doi/10.1126/science.3672124 (visited on 11/16/2022).

[175] Weisz, N. et al. “Oscillatory Alpha Modulations in Right Auditory Regions Reflect the Validity of Acoustic Cues in an Auditory Spatial Attention Task”. en. In: *Cerebral Cortex* 24.10 (Oct. 2014), pp. 2579–2590. ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bht113. URL: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bht113 (visited on 10/14/2022).

[176] Whitton, J. L., Lue, F., and Moldofsky, H. “A spectral method for removing eye movement artifacts from the EEG”. en. In: *Electroencephalography and Clinical Neurophysiology* 44.6 (June 1978), pp. 735–741. ISSN: 00134694. DOI: 10.1016/0013-4694(78)90208-0. URL: https://linkinghub.elsevier.com/retrieve/pii/0013469478902080 (visited on 10/04/2022).

[177] Womelsdorf, T. and Fries, P. “The role of neuronal synchronization in selective attention”. en. In: *Current Opinion in Neurobiology* 17.2 (Apr. 2007), pp. 154–160. ISSN: 09594388. DOI: 10.1016/j.conb.2007.02.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959438807000268 (visited on 10/14/2022).

[178] Womelsdorf, T., Schoffelen, J.-M., et al. “Modulation of Neuronal Interactions Through Neuronal Synchronization”. en. In: *Science* 316.5831 (June 2007), pp. 1609–1612. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1139597. URL: https://www.science.org/doi/10.1126/science.1139597 (visited on 11/17/2022).

[179] Wong, D. D. E. et al. “A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding”. In: *Frontiers in Neuroscience* 12 (Aug. 2018), p. 531. ISSN: 1662-453X. DOI: 10.3389/fnins.2018.00531. URL: https://www.frontiersin.org/article/10.3389/fnins.2018.00531/full (visited on 07/07/2022).

[180] Worden, M. S. et al. “Anticipatory Biasing of Visuospatial Attention Indexed by Retinotopically Specific α-Bank Electroencephalography Increases over Occipital Cortex”. en. In: *The Journal of Neuroscience* 20.6 (Mar. 2000), RC63–RC63. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.20-06-j0002.2000. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.20-06-j0002.2000 (visited on 10/14/2022).

[181] Wöstmann, M. et al. “Spatiotemporal dynamics of auditory attention synchronize with speech”. en. In: *Proceedings of the National Academy of Sciences* 113.14 (Apr. 2016), pp. 3873–3878. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1523357113. URL: https://pnas.org/doi/full/10.1073/pnas.1523357113 (visited on 10/14/2022).

[182] Wu, C.-T. et al. “The neural circuitry underlying the executive control of auditory spatial attention”. en. In: *Brain Research* 1134 (Feb. 2007), pp. 187–198. ISSN: 00068993. DOI: 10.1016/j.brainres.2006.11.088. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006899306034627 (visited on 07/07/2022).

[183] Xing, D. et al. “Laminar analysis of visually evoked activity in the primary visual cortex”. en. In: *Proceedings of the National Academy of Sciences* 109.34 (Aug. 2012), pp. 13871–13876. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1201478109. URL: https://pnas.org/doi/full/10.1073/pnas.1201478109 (visited on 10/20/2022).

[184] Xu, B. et al. “Improving Classification by Feature Discretization and Optimization for fNIRS-based BCI.” In: *Journal of Biomimetics Biomaterials and Tissue Engineering* 19.1 (Dec. 2013). DOI: 10.4172/1662-100X.1000119.

[185] Yagi, A. “Visual signal detection and lambda responses”. en. In: *Electroencephalography and Clinical Neurophysiology* 52.6 (Dec. 1981), pp. 604–610. ISSN: 00134694. DOI: 10.1016/0013-4694(81)91434-6. URL: https://linkinghub.elsevier.com/retrieve/pii/0013469481914346 (visited on 10/28/2022).

[186] Yamagishi, N. et al. “Attentional modulation of oscillatory activity in human visual cortex”. en. In: *NeuroImage* 20.1 (Sept. 2003), pp. 98–113. ISSN: 10538119. DOI: 10.1016/S1053-8119(03)00341-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811903003410 (visited on 10/14/2022).

[187] Ye, J. et al. “NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy”. en. In: *NeuroImage* 44.2 (Jan. 2009), pp. 428–447. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.08.036. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811908009695 (visited on 08/08/2022).

[188] Yücel, M. A. et al. “Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy”. en. In: *Journal of Innovative Optical Health Sciences* 07.02 (Mar. 2014), p. 1350066. ISSN: 1793-5458, 1793-7205. DOI: 10.1142/S1793545813500661. URL: https://www.worldscientific.com/doi/abs/10.1142/S1793545813500661 (visited on 07/07/2022).

[189] Zhang, M., Alamatsaz, N., and Ihlefeld, A. “Hemodynamic Responses Link Individual Differences in Informational Masking to the Vicinity of Superior Temporal Gyrus”. In: *Frontiers in Neuroscience* 15 (July 2021), p. 675326. ISSN: 1662-453X. DOI: 10.3389/fnins.2021.675326. URL: https://www.frontiersin.org/articles/10.3389/fnins.2021.675326/full (visited on 07/07/2022).

[190] Zhang, M., Mary Ying, Y.-L., and Ihlefeld, A. “Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy”. en. In: *Trends in Hearing* 22 (Jan. 2018), p. 233121651881746. ISSN: 2331-2165, 2331-2165. DOI: 10.1177/2331216518817464. URL: http://journals.sagepub.com/doi/10.1177/2331216518817464 (visited on 07/07/2022).

[191] Zwaag, W. van der et al. “Where sound position influences sound object representations: A 7-T fMRI study”. en. In: *NeuroImage* 54.3 (Feb. 2011), pp. 1803–1811. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.10.032. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811910013248 (visited on 07/07/2022).

CURRICULUM VITAE