Optimized Decoding for Auditory Attention Detection

Bradley Robert Ekin

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2016

Supervisory Committee:

Les Atlas

Adrian KC Lee

Chet Moritz

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Optimized Decoding for Auditory Attention Detection

Bradley Robert Ekin

Chair of the Supervisory Committee:
Professor Les Atlas
Electrical Engineering

The method of *stimulus reconstruction* has shown to be an effective tool for detecting a listener's attentional focus in a multi-talker environment. Using electroencephalography (EEG), this technique aims to learn neural decoding functions to predict a signal which is most similar to the temporal amplitude envelope of an attended talker's speech. By comparing this prediction to the envelope of each speech source in the environment, a decision can be made as to which source the listener is attending to. However, the conventional method for stimulus reconstruction is incomplete when applied to multi-talker environments. This is because the standard minimum mean square error (MMSE) criterion used for learning neural decoder functions discards information relating to how the brain jointly encodes both attended and unattended speech stimuli, discarding information which could be used for developing more discriminative decoders for auditory attention detection.

This thesis proposes how the conventional method of stimulus reconstruction can be improved by incorporating concepts from linear discriminant analysis (LDA). Utilizing the expected neural encoding properties to all attentional stimuli, we show how reconstruction error can be minimized while simultaneously maximizing the distance between the attentional class similarity metrics used for attention detection. This thesis then proposes how the method of stimulus reconstruction can be performed using only the spatial component of the neural response, improving computational efficiency by significantly reducing the number of

neural features used for attention detection. By employing the utility of canonical correlation analysis (CCA) to relate this spatial neural response to a temporal window of stimulus lags, we show how detection accuracy comparable to traditional stimulus reconstruction can be achieved; accuracies which further improve by adapting concepts from LDA into this reduced-rank framework for auditory attention detection.

TABLE OF CONTENTS

Page

List of Figures ii

List of Tables iv

Chapter 1: Introduction 1
1.1 Background 1
1.2 Overview of Thesis 5

Chapter 2: Stimulus Reconstruction for Auditory Attention Detection 9
2.1 Introduction 9
2.2 Stimulus Reconstruction 10
2.3 Stimuli-Difference Decoding 11

Chapter 3: Low-Rank Decoding for Auditory Attention Detection 15
3.1 Introduction 15
3.2 Multimodal Subspace Decoding 16
3.3 Canonical Correlation Analysis 17
3.4 Stimuli-Difference Subspace Decoding 19

Chapter 4: Results 21
4.1 Experimental Setup 21
4.2 Detection and Decoding Accuracy 23

Chapter 5: Conclusion 30

Bibliography 33

LIST OF FIGURES

Figure Number Page

2.1 Comparison of stimulus reconstruction methods where the dotted line is the detection threshold for SRAAD. **Left:** maximizing correlation of the reconstruction with the true attended stimulus without accounting for the unattended stimulus within the reconstruction method, i.e. minimize $\mathbb{E}\{|\boldsymbol{d}^T\boldsymbol{r} - s_{att}|^2\}$. **Right:** maximizing the correlation separation between $c_{\hat{a},a}$ and $c_{\hat{a},u}$, where $c_{\hat{a},a}$ is positive, for optimal reconstruction orthogonal to the detection threshold. 12

4.1 Subject specific detection accuracy for each decoding method. The dotted line indicates the classification performance level at which detection-accuracy is significantly greater than chance (63.33%) based on a binomial test at the 5% significance level. **Top**: detection accuracy of an attended speech stimulus. **Bottom**: detection accuracy of an unattended speech stimulus. $p > 5 \times 10^{-2}$ (*), $p < 5 \times 10^{-2}$ (**), $p < 5 \times 10^{-3}$ (***), $p < 5 \times 10^{-4}$ (*** *), $p < 5 \times 10^{-5}$ (*** **), $p < 5 \times 10^{-6}$ (*** ***). 24

4.2 Decoding accuracy for each method discussed. Each plot details the average sample cross-correlation of the decoded neural response with the true speech envelope (or projected window of the speech envelope for CCA and SDCCA). Correlations are averaged across all subjects and trials. Sample cross-correlation is defined as $c_{\hat{x},y}(\tau) = \text{corr}\{\hat{s}_x(t), s_y(t-\tau)\}$ or $c^{sub}_{\hat{x},y}(\tau) = \text{corr}\{\hat{\psi}_x(t), \psi_y(t-\tau)\}$, where $\text{corr}\{\cdot\}$ is the sample Pearson correlation coefficient of equation 2.6. **Top row**: decoding accuracy for predicting an attended speech envelope. **Bottom row**: decoding accuracy for predicting an unattended speech envelope. 26

4.3 The mean correlation coefficients obtained for each subject. The red line indicates direction and magnitude of the trend based on mean correlation over all subjects and trials. **Top row**: correlation coefficients obtained when predicting an attended speech envelope. **Bottom row**: correlation coefficients obtained when predicting an unattended speech envelope. 28

4.4 Comparison of each proposed decoding method against traditional MMSE estimation. Blue dots denote the mean difference between detection statistic correlations obtained for each subject. The red dot indicates mean correlation difference over all subjects and trials. **Top row**: correlation difference obtained when predicting an attended speech envelope. **Bottom row**: correlation difference obtained when predicting an unattended speech envelope. . 29

LIST OF TABLES

Table Number Page

4.1 Average detection accuracy ± s.d. across all subjects for detecting (**left**) an attended speech stimulus and (**right**) an unattended speech stimulus. 23

4.2 Average decoding accuracy ± s.d. across all trials and subjects. (·) denotes the difference between the correlation of the predicted stimulus with the true stimulus and the correlation of the prediction with the opposite stimulus (see text). 27

ACKNOWLEDGMENTS

To my friends, family, and colleagues who have helped inspire and encourage me to pursue my dreams, I thank each and every one of you for your unwavering support and friendship; I would not be where I am today without you. Special thanks to my advisor and mentor, Prof. Les Atlas, for your guidance, inspirational advice, and for providing me with exciting and new directions to explore my scientific interests. Thank you to Prof. Adrian KC Lee for your vast knowledge in neuroscience and continued support in my research. Thank you to Prof. Chet Moritz for your time, your invaluable advice, and for kindly agreeing to be a part of my research committee. Thanks also to Scott Wisdom, Tyler Ganter, Tommy Powers, Majid Mirbagheri, Eldridge Alcantara and Dave Dolengewicz for your insightful discussions and help over the years, and most importantly, thank you for your amazing friendship. Thank you to my family, especially my parents and brother Trevor, for providing a lifetime of continuous support, encouragement, wisdom, and love. And most of all, thank you to my wonderful wife Linnea, who motivates me and makes every day as wonderful as the last. Your love and friendship keeps me going.

DEDICATION

To my mom, dad, brother Trevor, and wife Linnea.

Chapter 1

INTRODUCTION

1.1 Background

Humans possess the remarkable ability to "tune in" to speech sources of interest in complex auditory environments. When this environment is comprised of other competing talkers, this innate human ability of selecting a speech source to "attend" to, while simultaneously filtering out distracting speech, is known as selective auditory attention. Coined *the cocktail party problem* by Colin Cherry [9], the neural underpinnings of how the human brain effortlessly solves this non-trivial task have been a key research topic for decades. As a result, the field has progressed to where it is now possible to detect who a listener is focusing their attention on by comparing their neural activity to the activity of multiple candidate speech sources in an environment [27]. This detection framework is commonly referred to as auditory attention detection (AAD).

The AAD paradigm has motivated many researchers and industry personnel due to its far-reaching number of applications. For example, AAD could provide additional clinical tools for the analysis and diagnosis of attention-related disorders. For more industry-driven approaches, AAD can be linked to providing new mediums for information retrieval within the advertising and marketing industries, as well as for other brain-computer interface (BCI) applications which could benefit from a user's auditory focus as input. For example, an application of AAD within the evolving field of virtual reality would be to incorporate neural feedback to govern how novel scenarios unfold in a controlled auditory and visual environment based on the subject's auditory focus. Another example, and one of the more common AAD applications within the literature, is incorporating a listener's auditory focus into the processing of hearing prostheses (such as hearing aids or cochlear implants). By integrating

neural feedback into hearing prostheses, researchers hope to construct systems which utilize acoustic beamforming techniques to steer nulls toward the angular directions of distracting speech sources while retaining and amplifying the speech of an attended source, as in the recent works of [10, 21, 35].

Almost all of these applications would have a common set of requirements: 1) provide minimal impact to the user's comfort level (for example, high-density neural recording caps would likely not be a viable option), 2) computationally efficient processing strategies (especially for battery-powered wearable technologies), and most importantly 3) robust performance using a noninvasive recording modality. Developing AAD strategies using noninvasive techniques, such as magnetoencephalography (MEG) and electroencephalography (EEG), has been at the forefront of this research. This includes methods like state-space modeling for tracking dynamic changes in attention switching for MEG [2, 3], empirical mode decomposition for modeling gamma band synchronization between attended stimuli and neural activity for EEG [20], automatic independent component analysis for attention detection for EEG [28], as well as a probabilistic approach based on hidden Markov model regression for combined M-EEG [25]. However, the technique which is most commonly used throughout the literature, and which is a central focus of this thesis, is AAD via stimulus reconstruction using EEG.

1.1.1 Stimulus reconstruction

The method of stimulus reconstruction is a dominant technique used in the study of neural processing. This technique has provided a clever way to infer extremely complex encoding properties within a population of neurons by learning linear mappings from the neural population back to naturally occurring stimuli. For example, stimulus reconstruction has shown how particular areas of the brain encode information by reconstructing estimations of visual scenes [6, 34, 36], as well as auditory scenes [24, 29, 31, 32], based on selective invasive recording locations within the brain. Another example is how the brain processes speech, where stimulus reconstruction has allowed for quantitative analysis on how the brain encodes

stimuli by comparing how well different stimulus representations account for observed neural responses [29]. This simple, yet effective assumption of a linear model has allowed stimulus reconstruction to become a powerful tool for neural encoding analysis.

The underlying model used for stimulus reconstruction assumes that a neural response can be represented by a particular stimulus feature convolved with a neural impulse response function. Most often, any nonlinearities within this relationship are absorbed within the particular stimulus feature being evaluated. For example, an important relationship commonly used within auditory neuroscience literature is that the cortical neural response follows the temporal amplitude envelope of speech [1, 18, 22]. This low-frequency relationship between neural activity and the speech envelope has allowed for less resolved noninvasive techniques like MEG and EEG to become practical tools for auditory analysis. These neural impulse responses are commonly referred to as temporal response functions (TRFs) and are modeled by:

$$r[c,n] = \sum_{\ell=0}^{L-1} h[c,\ell]s[n-\ell] + \eta[c,n], \tag{1.1}$$

where $r[c,n]$ denotes the discrete-time neural response observed at the c^{th} measurement channel, $h[c,\ell]$ denotes the respective TRF where ℓ denotes discrete lag index, $s[n]$ denotes the speech envelope, and $\eta[c,n]$ denotes residual noise not explained by the model and is assumed to be Gaussian. Under this notation, stimulus reconstruction attempts to deconvolve the TRF via a neural decoding function $d[c,\ell]$, where weighted and delayed neural observations are combined such that

$$\hat{s}[n] = \sum_{c=1}^{C}\sum_{\ell=0}^{L-1} d[c,\ell]r[c,n+\ell] \tag{1.2}$$

results in a prediction of the stimulus envelope. These decoders are typically learned using minimum-mean square error (MMSE) optimization criterion, as follows:

$$\text{minimize } \mathbb{E}\{|\hat{s}-s|^2\}, \tag{1.3}$$

where $\mathbb{E}\{\cdot\}$ denotes expected value.

Stimulus reconstruction has also proven to be an insightful technique for the study of selective auditory attention, where it has shown to be sensitive to cocktail party paradigms [11, 12, 15, 23, 30]. More specifically, these studies have shown overwhelming evidence that both attended and unattended speech stimuli are simultaneously encoded within the brain, where attention is controlled by a combination of bottom-up and top-down cognitive processing. Under a simple two-talker scenario, this follows the model of:

$$r[c,n] = \sum_{\ell=0}^{L-1} h_{att}[c,\ell]s_{att}[n-\ell] + h_{unt}[c,\ell]s_{unt}[n-\ell] + \eta[c,n], \tag{1.4}$$

where $h_{att}[c,\ell]$ denotes the attended TRF to an attended speech envelope $s_{att}[n]$, and $h_{unt}[c,\ell]$ denotes the unattended TRF to an unattended speech envelope $s_{unt}[n]$. Because the encoding properties for both TRFs are separable [11], predictions of either the attended or unattended speech source can be recovered by regressing the decoder in equation 1.2 toward the desired stimulus (i.e. replacing s in equation 1.3 with either s_{att} or s_{unt}). This, in turn, has allowed stimulus reconstruction to become an effective tool for AAD.

1.1.2 Auditory attention detection via stimulus reconstruction

Auditory attention detection by means of stimulus reconstruction (SRAAD), as proposed by O'Sullivan *et al.*, 2014 [27], is a technique which attempts to reconstruct an estimation of the attended speech envelope using the expected neural encoding properties for an attended stimulus (learned by equation 1.3 when $s = s_{att}$). Then, given numerous talkers in an environment, AAD is performed by determining which of the speech sources is most similar to the reconstruction. Using a sample Pearson correlation coefficient as the similarity measure, this method has been shown to perform well above chance for detecting a listener's attentional focus when using their EEG response.

In addition, this technique has allowed for each of the concerns for practical AAD applications regarding user comfort, computational/power efficiency, and robust noninvasive detection accuracy to be addressed. For example, [26] has shown that SRAAD is robust to the number of EEG channels used for stimulus reconstruction, suggesting that fine-grained

spatial sampling is not necessary for attention detection, addressing both user comfort and power efficiency. As for robust detection, [7] proposed how SRAAD accuracy can be improved by using speech envelope extraction techniques representative to auditory pathway models, allowing a more biologically-inspired speech envelope to relate to the EEG. In another study, [4] suggested that SRAAD is robust under the assumption that access to the clean and separate speech sources is not available for reconstruction comparison (representing the blind source separation stage for a hearing prosthesis). SRAAD was also shown to be robust to attentional task in [19], where attentional detection in a dichotic speech paradigm was accurately decoded under numerous low-level to high-level task conditions.

All of these studies rely on a common stimulus reconstruction method: minimizing the mean square error between the predicted stimulus and the true stimulus following that of equation 1.3. However, this type of stimulus reconstruction does not take into account any expected encoding properties of the opposite stimulus. That is, when regressing decoders for the reconstruction of an attended stimulus, no knowledge of the unattended stimulus is accounted for within the optimization criterion. This poses a problem, especially when the detection statistic for SRAAD relies on comparing the similarity between the reconstruction with both true attended and unattended stimuli. By not utilizing all of the expected attentional encoding properties within the stimulus reconstruction technique, useful information needed for attention detection is being discarded; information which could provide significant impact to the SRAAD framework if taken advantage of in the correct way.

1.2 Overview of Thesis

This thesis addresses the lack of accountability for all attentional stimuli assumed within the model so that the expected neural encoding properties to both attended and unattended stimuli are fully utilized. This is achieved by proposing an optimization criterion based on linear discriminant analysis (LDA), where the goal is to minimize reconstruction error while maximizing the attentional class separation of correlation coefficients used for the detection statistic of SRAAD. More specifically, we aim to maximize the separation between how well

the reconstruction correlates with a true attended speech envelope, compared to its correlation with a true unattended speech envelope. It is shown how this optimization criterion directly maximizes the detection statistic for SRAAD and how minimal modifications to traditional MMSE estimation can be made which result in significant detection accuracy improvements. The proposed framework is evaluated using experimental data from 34 EEG subjects who participated in a two-talker selective auditory attention task (following the model of equation 1.4). It is shown that, on average, detection accuracy improved compared to traditional stimulus reconstruction via MMSE estimation by +5.6% (absolute) for detection of an attended stimulus and +13.6% (absolute) for detection of an unattended stimulus.

Additionally, this thesis addresses the issue of computational efficiency for practical SRAAD applications. Current methods for stimulus reconstruction relate high-dimensional spatiotemporal activity of the neural response to that of a single instance of the speech stimulus. In many cases, this results in ill-conditioned covariance estimators where careful processing strategies are required for numerically stable decoder function estimations, such as applying principal component analysis (PCA) or regularization. This thesis instead proposes how the spatiotemporal relationship between the neural responses and stimuli can be split between these modalities, where the spatial component of EEG is related to the temporal statistics of the speech envelope. This strategy reduces the total number of features required for stimulus prediction and allows for more numerically stable processing techniques to be applied. Referred to as generalized reconstruction for auditory attention detection (GRAAD), it is shown how attention detection can be performed within this reduced-rank framework, where canonical correlation analysis (CCA) can be applied to find a maximally correlated subspace between these two modalities. It is shown that this alternative approach results in comparable detection accuracies to that of traditional stimulus reconstruction via MMSE estimation, even though less features are used within the analysis.

Finally, it is shown how the proposed GRAAD framework can be modified to include information relating to all attentional stimuli assumed within the model to improve detection

accuracy. Similar to the approach based on LDA for SRAAD, here we present an optimization criterion which finds a subspace between the EEG and stimulus modalities which maximizes the attentional class separation of correlation coefficients used for the detection statistic of GRAAD. It is shown that this proposed framework results in the highest average improvement for detection accuracy when compared to traditional stimulus reconstruction via MMSE estimation, improving by +6.4% (absolute) for detection of an attended stimulus and +14.4% (absolute) for detection of an unattended stimulus.

1.2.1 Thesis contents

This thesis is made up of four parts. The first part, outlined in chapter 2, focuses on stimulus reconstruction techniques for SRAAD. The general framework needed for SRAAD, which includes defining the detection statistic as well as how neural decoding is performed, is outlined in section 2.1. Section 2.2 introduces the standard technique for stimulus reconstruction where neural decoders are trained using MMSE criterion between the prediction and the desired stimulus in which to reconstruct. In section 2.3, stimulus reconstruction framework is proposed based on concepts from LDA, where it is shown how standard MMSE estimation can be modified to account for all attentional stimuli assumed within the model.

The second part of this thesis, outlined in chapter 3, introduces the GRAAD framework as a reduced-rank alternative to SRAAD. The motivation for this type of approach is addressed in section 3.1. The foundation for how the spatiotemporal relationship between the EEG and stimulus can be represented as a subspace projection between these modalities, as well as how auditory attention detection can be performed, is developed in section 3.2. Next, section 3.3 derives how CCA can be used for learning optimal projections for both of these modalities. Finally, section 3.4 shows how the LDA based concepts proposed in section 2.3 can be adapted to the GRAAD framework for improving detection accuracy.

The third part of this thesis is outlined in chapter 4, where each of the AAD techniques presented in chapters 2 and 3 are evaluated and compared using real EEG data. The experimental setup which includes participant information, stimuli and procedure setup, data

acquisition, and preprocessing techniques is detailed in section 4.1. Evaluation and comparison for each of the methods is discussed in section 4.2.

Chapter 5, the final part of this thesis, summarizes the methods proposed and suggests promising directions for future work.

1.2.2 Summary of Notation

The discrete-time speech envelope of an attended talker is denoted by $s_{att}[n]$, and unattended talker denoted by $s_{unt}[n]$ for $n = 0, 1, \ldots, N-1$. The EEG response of the listener is denoted by $r[c, n+\ell]$, where $c = 1, 2, \ldots, C$ denotes the EEG electrode channel and $\ell = 0, 1, \ldots, L-1$ denotes the lag index. All signals are assumed to have been centered to zero mean. $\mathbb{E}\{\cdot\}$ denotes statistical expectation. All matrices are denoted by a bold uppercase letter, e.g. $\mathbf{H}$, and vectors as bold lowercase italic letters, e.g. $\boldsymbol{x}$. A vector or matrix transpose is denoted by $(\cdot)^T$. Column vectors are assumed unless otherwise stated.

Chapter 2

STIMULUS RECONSTRUCTION FOR AUDITORY ATTENTION DETECTION

2.1 Introduction

The method of stimulus reconstruction used within the SRAAD framework is performed by sampling, weighting, and combining time-lagged copies of each EEG channel together so that a prediction of the underlying stimulus is recovered. These filter weights are commonly referred to as a neural decoder function and take the form of:

$$\boldsymbol{d} = \Big[d[1,0] \quad d[1,1] \quad \cdots \quad d[1,L-1] \quad d[2,0] \quad \cdots \quad d[C,L-1] \Big]^T. \tag{2.1}$$

By defining the respective window for the EEG response as:

$$\boldsymbol{r}[n] = \Big[r[1,n] \quad r[1,n+1] \quad \cdots \quad r[1,n+L-1] \quad r[2,n] \quad \cdots \quad r[C,n+L-1] \Big]^T, \tag{2.2}$$

the decoding functionality of equation 1.2 can be represented as the inner product between the decoder and neural response, defined by equation 2.3 below.

$$\hat{s}[n] = \boldsymbol{d}^T \boldsymbol{r}[n] \tag{2.3}$$

Decoding functions are typically trained and tested using a leave-one-out cross-validation approach across neural responses obtained from K similar length trials. Typical experiments for AAD limit the total number of stimuli to only two simultaneously spoken narratives (each trial of approximately 60 seconds in length), following the model of equation 1.4. Reconstruction similarity is tested by measuring how well the predicted speech envelope correlates with the true speech envelope for both the attended and unattended stimuli. Assuming the stimulus being predicted is the attended speech source, let these sample correlation coefficients

for the k^{th} test trial be denoted as:

$$\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}[k] = \mathrm{corr}\{\hat{s}_{att}[k,n], s_{att}[k,n]\} \tag{2.4}$$

$$\mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}[k] = \mathrm{corr}\{\hat{s}_{att}[k,n], s_{unt}[k,n]\}, \tag{2.5}$$

where

$$\mathrm{corr}\{x[n], y[n]\} = \frac{\sum_n (x[n]-\bar{x})(y[n]-\bar{y})}{\sqrt{\sum_n (x[n]-\bar{x})^2}\sqrt{\sum_n (y[n]-\bar{y})^2}} \tag{2.6}$$

is the sample Pearson correlation coefficient[1], and $\bar{x}$ and $\bar{y}$ denote the sample means. Provided we know ground truth for which speech source the listener is attending to, detection accuracy for SRAAD is measured by comparing these correlations via an indicator function $I_{att}[k]$ which determines whether the k^{th} test trial detected the attended stimulus correctly or not. This function is shown by equation 2.7 below.

$$I_{att}[k] = \begin{cases} 1 & \text{if } \mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}[k] > \mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}[k] \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

Detection accuracy for an attended stimulus is defined as the average value of I_{att}. Detection of an unattended stimulus is performed analogously. With both the detection statistic and decoding functionality defined, we can now develop optimization criteria for the learning of decoder weights.

2.2 Stimulus Reconstruction

The traditional method of stimulus reconstruction when applied to AAD aims to learn decoder weights which minimize the mean square error between $\hat{s}$ and s_{att}, following the MMSE cost function of:

$$\boldsymbol{d}_{MMSE} = \arg\min_{\boldsymbol{d}} \frac{1}{2}\mathbb{E}\left\{\left|\boldsymbol{d}^T\boldsymbol{r} - s_{att}\right|^2\right\}. \tag{2.8}$$

[1]The sample Pearson correlation coefficient is a similarity measure between two sampled random variables x and y, where the correlation coefficient is bounded between $-1 \leq \mathrm{c}_{\mathrm{x,y}} \leq 1$. Perfect correlation occurs when $x = y$ which results in $\mathrm{c}_{\mathrm{x,y}} = 1$, and perfect anti-correlation occurs when $x = -y$ which results in $\mathrm{c}_{\mathrm{x,y}} = -1$.

The optimal solution to this quadratic problem is the least squares solution of:

$$\boldsymbol{d}_{MMSE} = \boldsymbol{\Sigma}_{rr}^{-1}\boldsymbol{\sigma}_{rs_a}, \tag{2.9}$$

where $\boldsymbol{\Sigma}_{rr}$ denotes the covariance matrix of $\boldsymbol{r}$, and $\boldsymbol{\sigma}_{rs_a}$ denotes the cross-covariance between $\boldsymbol{r}$ and s_{att}.

MMSE criterion is advantageous for SRAAD as it can be reformulated to the correlation maximization problem of:

$$\arg\max_{\boldsymbol{d}} \quad \frac{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}s_{att}\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{d}\right\}}\sqrt{\mathbb{E}\left\{s_{att}^2\right\}}}. \tag{2.10}$$

Due to a detection statistic which relies on how well the reconstruction correlates with the true stimulus, showing that $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}$ is maximized within this reconstruction technique is an important characteristic to have.

Stimulus reconstruction via MMSE estimation is popular among the SRAAD community due to its simplicity and reasonable prediction accuracy. However, this reconstruction technique does not account for the joint encoding properties of both the attended and unattended speech stimuli within the neural population being measured. Because of this, information relating to the unattended stimulus is discarded within the regression technique, effectively ignoring half of the detection statistic for SRAAD (i.e. $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}$). To see how this can be addressed, let us consider a cost function which accounts for all stimuli assumed within the selective attention model.

2.3 Stimuli-Difference Decoding

Reconsidering the detection statistic of equation 2.7, the goal is to find a neural decoding function which maximizes the correlation of the reconstruction with an attended speech envelope. Furthermore, the correlation of the reconstruction with an unattended speech envelope should be minimized. Noting that correlation is bounded between $-1 \leq \mathrm{c}_{\mathrm{x,y}} \leq 1$, the optimal decoder for this detection statistic relates to maximizing the separation between $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}$ and $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}$ (i.e. forcing $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}} \to 1$ and $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}} \to -1$). To achieve this, the reconstruction

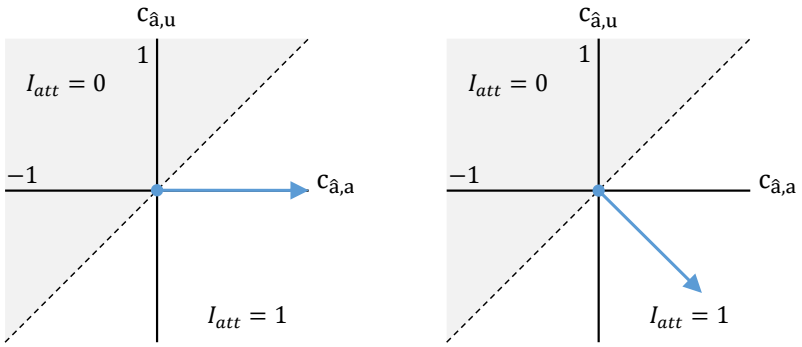

Figure 2.1: Comparison of stimulus reconstruction methods where the dotted line is the detection threshold for SRAAD. **Left:** maximizing correlation of the reconstruction with the true attended stimulus without accounting for the unattended stimulus within the reconstruction method, i.e. minimize $\mathbb{E}\{|\boldsymbol{d}^T\boldsymbol{r} - s_{att}|^2\}$. **Right:** maximizing the correlation separation between $c_{\hat{a},a}$ and $c_{\hat{a},u}$, where $c_{\hat{a},a}$ is positive, for optimal reconstruction orthogonal to the detection threshold.

technique must embed both attentional stimuli into the reconstruction, where $c_{\hat{a},a} - c_{\hat{a},u}$ is maximized. Similar to LDA, this reconstruction technique looks to find a decoder which minimizes the reconstruction error while maximizing separation between the two attentional classes within the detection statistic. This is achieved by decoding $\boldsymbol{r}$ to a point of correlation with each stimulus which is orthogonal to the detection threshold (see figure 2.1).

To convey this mathematically, consider the following correlation maximization cost function.

$$\arg\max_{\boldsymbol{d}} \quad \frac{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}s_{att}\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{d}\right\}}\sqrt{\mathbb{E}\left\{s_{att}^2\right\}}} - \frac{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}s_{unt}\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{d}\right\}}\sqrt{\mathbb{E}\left\{s_{unt}^2\right\}}} \tag{2.11}$$

Solving this optimization problem yields the optimal reconstruction statistics required for attention detection, where $c_{\hat{a},a} > c_{\hat{a},u}$. By normalizing all stimuli to have unit variance, i.e. $\mathbb{E}\{s_{att}^2\} = \mathbb{E}\{s_{unt}^2\} = 1$ (assuming variables have been centered), the cost function of equation 2.11 reduces to:

$$\arg\max_{\boldsymbol{d}} \quad \frac{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}(s_{att} - s_{unt})\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{d}\right\}}}. \tag{2.12}$$

Interestingly enough, and again noting that correlation maximization can be reformulated

as MMSE estimation, this optimization problem can now be posed as minimizing the mean square error between the prediction and the difference between both normalized attentional stimuli.

$$\arg\min_{\boldsymbol{d}} \frac{1}{2}\mathbb{E}\left\{\left|\boldsymbol{d}^T\boldsymbol{r} - (s_{att} - s_{unt})\right|^2\right\} \tag{2.13}$$

Here, s_{att} and s_{unt} are assumed to be independent zero-mean random variables, where $\mathbb{E}\{s_{att}s_{unt}\} = 0$, which is a safe assumption due to the independence of the talkers. The closed form solution to this quadratic problem is similar to that of LDA[2], where the optimal decoder is defined as:

$$\boldsymbol{d}_{SDMMSE} = \boldsymbol{\Sigma}_{rr}^{-1}(\boldsymbol{\sigma}_{rs_a} - \boldsymbol{\sigma}_{rs_u}) = \boldsymbol{\Sigma}_{rr}^{-1}\boldsymbol{\sigma}_{rs_{au}}, \tag{2.14}$$

where $\boldsymbol{\sigma}_{rs_{au}}$ denotes the cross-covariance between the neural response $\boldsymbol{r}$ and stimuli difference $s_{att} - s_{unt}$. We refer to this decoding technique as stimuli-difference minimum-mean square error (SDMMSE) estimation.

Because the SDMMSE cost function aims to find decoders which produce correlations orthogonal to the detection threshold line, a decoder trained to detect the attended stimulus can also be used to detect the unattended stimulus simply by decoder negation (i.e. $\boldsymbol{d}^{att}_{SDMMSE} = -\boldsymbol{d}^{unt}_{SDMMSE}$). This provides a sort of "one-fits-all" decoding scheme, whereas the minimizer of $\mathbb{E}\{|\boldsymbol{d}^T\boldsymbol{r} - s_{unt}|^2\}$ would be required for traditional MMSE based stimulus reconstruction. Another attractive feature with this setup is that it easily generalizes to a model with multiple unattended talkers. Assuming a total of N_s speech sources, where one is attended to and the j^{th} unattended talker is denoted as $s^{(j)}_{unt}$, the stimuli-difference reconstruction framework can be generalized by solving:

$$\boldsymbol{d}_{SDMMSE} = \arg\min_{\boldsymbol{d}} \frac{1}{2}\mathbb{E}\left\{\left|\boldsymbol{d}^T\boldsymbol{r} - \left(s_{att} - \sum_{j=1}^{N_s-1} s^{(j)}_{unt}\right)\right|^2\right\}, \tag{2.15}$$

[2]The homoscedasticity assumption for LDA is that the attentional class covariance's are identical and equal to the covariance of the EEG response (i.e. $\mathbb{E}\{(\boldsymbol{r}s_{att})^2\} = \mathbb{E}\{(\boldsymbol{r}s_{unt})^2\} = \mathbb{E}\{\boldsymbol{r}^2\}$). This assumption relaxes the within class minimization typical to that of Fisher's linear discriminant.

assuming all stimuli are uncorrelated and have been normalized. This addresses a model which is more true to the classic cocktail party scenario. However, due to the lack of experimental data within this context, this thesis will not address empirical performance measures under a multi-unattended talker model, nor do we suggest the brain encodes multiple unattended stimuli as this linear sum for large N_s (see, for example [33]).

Chapter 3

LOW-RANK DECODING FOR AUDITORY ATTENTION DETECTION

3.1 Introduction

Up to this point, this thesis has focused on developing backward mapping techniques from a multi-channel neural response in time to a one-dimensional stimulus. That is, this concept requires a size $\mathbb{R}^{LC}$ vector of spatiotemporal neural responses to be decoded (projected) to $\mathbb{R}^1$. This chapter focuses on generalizing the spatiotemporal relationship between the neural response and speech stimulus modalities, reducing the overall number of features required for auditory attention detection and improving computational efficiency.

Currently, all spatiotemporal relationships between neural activity and stimuli are captured within the EEG channel observation windows of $\boldsymbol{r} \in \mathbb{R}^{LC}$. If we instead only consider the spatial component of the neural response, all temporal relationships must be absorbed into an observation window of stimulus lags. This concept effectively reduces the total number of neural observation features to $\mathbb{R}^C$, where the stimulus is now represented by a size $\mathbb{R}^L$ vector of causal time lags. This chapter introduces how canonical correlation analysis (CCA) can be used to find optimal projections which map both the spatial neural response and stimulus window vectors onto a common one-dimensional and maximally correlated subspace, resulting in a low-rank alternative to SRAAD. By reducing the total features used for analysis from LC in standard SRAAD techniques to now $L + C$, a slight loss in detection accuracy would be expected. However, we will show how the concepts of LDA which were proposed for SDMMSE can be adapted into this multimodal subspace framework to counteract this problem.

To motivate this type of reduced-rank analysis, consider a typical EEG setup where

$C = 128$ and a lag window of $L = 17$ samples is evaluated over (250 ms at $f_s = 64$ Hz). Traditional stimulus reconstruction framework would require an EEG covariance of size $\mathbf{\Sigma}_{rr} \in \mathbb{R}^{2,176\times 2,176}$ to be estimated. This is likely to result in an ill-conditioned estimator where careful processing strategies would be required for numerically stable inversions, such as PCA (e.g. [13]) or regularization (e.g. [4, 26]). This is even more evident when averaging decoders together which are trained on a trial-by-trial basis, where the trial length in many cases is less than twice the size of $\mathbf{\Sigma}_{rr}$. In the proposed method, sample covariance matrices of size $\mathbb{R}^{130\times 130}$ and $\mathbb{R}^{17\times 17}$ would be required for both the spatial neural response and stimulus lag window, respectively. We found that this resulted in better behaved computational steps when learning these multimodal projections and produced predictions which correlated more with the stimuli than with standard SRAAD techniques (see chapter 4 figure 4.2). Additionally, this type of processing strategy effectively reduces the number of multiplications needed for stimulus prediction by a factor of $\frac{N_s L+C}{LC}$, assuming $N_s < \frac{C(L-1)}{L}$, compared to conventional MMSE stimulus reconstruction (a factor of ≈ 0.074 when $N_s = 2$, and equality when $N_s \approx 120$), which could be an important savings when optimizing computational efficiency onboard battery-powered wearable technology.

3.2 Multimodal Subspace Decoding

Projecting both modalities onto a common subspace is proposed by relating the instantaneous neural responses observed at each EEG channel to a causal lag window of the speech envelope. Let the spatial neural response be defined as:

$$\boldsymbol{r}_c[n] = \begin{bmatrix} r[1,n] & r[2,n] & \cdots & r[C,n] \end{bmatrix}^T, \tag{3.1}$$

and the respective window of causal lags for the speech envelope be defined as:

$$\boldsymbol{w}[n] = \begin{bmatrix} s[n] & s[n-1] & \cdots & s[n-L+1] \end{bmatrix}^T. \tag{3.2}$$

Subspace decoding is then performed by finding projections which map both $\boldsymbol{r}_c$ and $\boldsymbol{w}$ onto a common one-dimensional space. In the context of auditory attention detection, the stimulus

projection must be applied to all stimuli assumed in the model. For the simple model of two competitive talkers, this relates to projecting both stimuli onto:

$$\psi_{att} = \boldsymbol{d}_w^T \boldsymbol{w}_{att} \tag{3.3}$$

$$\psi_{unt} = \boldsymbol{d}_w^T \boldsymbol{w}_{unt}. \tag{3.4}$$

The goal for subspace AAD is to now predict ψ_{att} from the spatial neural response, denoted by:

$$\widehat{\psi}_{att} = \boldsymbol{d}_r^T \boldsymbol{r}_c. \tag{3.5}$$

Revising the similarity measure used for SRAAD, attention detection for the k^{th} test trial is then performed by comparing the sample Pearson correlation coefficients between the projected neural response and projected stimuli, as follows:

$$\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}^{sub}[k] = \mathrm{corr}\{\widehat{\psi}_{att}[k,n], \psi_{att}[k,n]\} \tag{3.6}$$

$$\mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}^{sub}[k] = \mathrm{corr}\{\widehat{\psi}_{att}[k,n], \psi_{unt}[k,n]\}, \tag{3.7}$$

where detection accuracy is defined as the average value of the indicator function below.

$$I_{att}^{sub}[k] = \begin{cases} 1 & \text{if } \mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}^{sub}[k] > \mathrm{c}_{\hat{\mathrm{a}},\mathrm{u}}^{sub}[k] \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

We refer to this type of multimodal subspace decoding framework as generalized reconstruction for auditory attention detection (GRAAD).

With the subspace detection statistic defined, optimization criteria for the learning of $\boldsymbol{d}_r$ and $\boldsymbol{d}_w$ can now be developed. These criteria must be able to learn projections which maximize the correlation within the common subspace of the projected spatial neural response and projected stimulus lag window, following that of the ubiquitous multimodal analysis tool of CCA.

3.3 Canonical Correlation Analysis

The objective function of CCA aims to find a pair of linear transformations on two random vectors such that one component within each set of transformed variables is maximally

correlated with a single component of the other set [5, 16, 17]. In the context of GRAAD, this relates to the following correlation maximization problem:

$$\underset{\boldsymbol{d}_r,\boldsymbol{d}_w}{\arg\max} \quad \frac{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{w}_{att}^T\boldsymbol{d}_w\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{r}_c^T\boldsymbol{d}_r\right\}}\sqrt{\mathbb{E}\left\{\boldsymbol{d}_w^T\boldsymbol{w}_{att}\boldsymbol{w}_{att}^T\boldsymbol{d}_w\right\}}}. \tag{3.9}$$

This criterion aims to find a maximally correlated subspace between both the spatial neural response and the temporal stimulus window. CCA also imposes constraints so that each vector projection has unit variance. This results in the traditional optimization problem for CCA:

$$\begin{aligned}\boldsymbol{d}_{r-CCA}, \boldsymbol{d}_{w-CCA} = \underset{\boldsymbol{d}_r,\boldsymbol{d}_w}{\arg\max} \quad & \mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{w}_{att}^T\boldsymbol{d}_w\right\} \\ \text{subject to } & \mathbb{E}\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{r}_c^T\boldsymbol{d}_r\} = 1 \\ & \mathbb{E}\{\boldsymbol{d}_w^T\boldsymbol{w}_{att}\boldsymbol{w}_{att}^T\boldsymbol{d}_w\} = 1.\end{aligned} \tag{3.10}$$

Let the auto-covariances of $\boldsymbol{r}_c$ and $\boldsymbol{w}_{att}$ be denoted as $\boldsymbol{\Sigma}_{r_cr_c}$ and $\boldsymbol{\Sigma}_{w_aw_a}$, respectively, and the size $C \times L$ cross-covariance between them be denoted by $\boldsymbol{\Sigma}_{r_cw_a}$. The multimodal projections can now be found by solving the following generalized eigenvalue problem (as shown in [5]).

$$\begin{bmatrix} \mathbf{0} & \boldsymbol{\Sigma}_{r_cw_a} \\ \boldsymbol{\Sigma}_{r_cw_a}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_r \\ \boldsymbol{d}_w \end{bmatrix} = \mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}^{sub} \begin{bmatrix} \boldsymbol{\Sigma}_{r_cr_c} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{w_aw_a} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_r \\ \boldsymbol{d}_w \end{bmatrix} \tag{3.11}$$

In this problem, the eigenvalue represents the correlation between the projected spatial neural response and the projected stimulus lag window, where these multimodal projections are the eigenvectors. Therefore, the eigenvector which corresponds to the largest value of $\mathrm{c}_{\hat{\mathrm{a}},\mathrm{a}}^{sub}$ is the solution for both $\boldsymbol{d}_{r-CCA}$ and $\boldsymbol{d}_{w-CCA}$.

CCA is a heavily utilized tool across multiple disciplines and can be found as built in functions in numerous computing platforms (e.g. MATLAB, R, and Python). All CCA computations evaluated in this thesis were performed using MATLAB's `canoncorr` function. This function, along with other platforms, is particularly advantageous as it bypasses the direct evaluation of equation 3.11 where sample covariance estimations are avoided. This

is achieved by utilizing an algorithm based on QR decomposition and singular value decomposition (SVD) which are processed directly on the data [8, 14], resulting in a fast, efficient, and more numerically stable way to compute CCA projections.

It should also be noted that solutions for CCA not only provide linear transformations for the maximally correlated subspace, but can also provide successive projections, each a maximally correlated subspace with an additional constraint that it is orthogonal to the last (with the maximum number of projections equal to $\min\{C, L\}$ assuming both $\boldsymbol{r}$ and $\boldsymbol{w}$ are full rank). We do not address this additional capability of CCA, but recognize that this information may be useful and plan to investigate this more deeply in the future.

3.4 Stimuli-Difference Subspace Decoding

Consistent with the general theme of this thesis, useful information relating to how both attended and unattended stimuli are jointly encoded within the measured neural population is discarded when not accounting for all attentional stimuli in the decoding technique. Similar to the concepts presented in section 2.3 for SDMMSE, the attentional class separation concepts of LDA can be adapted into this reduced-rank framework as well. That is, the goal for GRAAD should aim to find projections which best separate $\mathrm{c}^{sub}_{\hat{\mathrm{a}},\mathrm{a}}$ and $\mathrm{c}^{sub}_{\hat{\mathrm{a}},\mathrm{u}}$ so that their joint correlation is orthogonal to the detection threshold line. This relates to solving the optimization problem of:

$$\underset{\boldsymbol{d}_r,\boldsymbol{d}_w}{\arg\max} \frac{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{w}_{att}^T\boldsymbol{d}_w\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{r}_c^T\boldsymbol{d}_r\right\}}\sqrt{\mathbb{E}\left\{\boldsymbol{d}_w^T\boldsymbol{w}_{att}\boldsymbol{w}_{att}^T\boldsymbol{d}_w\right\}}} - \frac{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{w}_{unt}^T\boldsymbol{d}_w\right\}}{\sqrt{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c\boldsymbol{r}_c^T\boldsymbol{d}_r\right\}}\sqrt{\mathbb{E}\left\{\boldsymbol{d}_w^T\boldsymbol{w}_{unt}\boldsymbol{w}_{unt}^T\boldsymbol{d}_w\right\}}}. \tag{3.12}$$

By making the assumption that both the attended and unattended stimulus lag windows have equal covariance, i.e. $\mathbb{E}\{\boldsymbol{w}_{att}\boldsymbol{w}_{att}^T\} = \mathbb{E}\{\boldsymbol{w}_{unt}\boldsymbol{w}_{unt}^T\} = \boldsymbol{\Sigma}_{ww}$, the objective function of equation 3.12 can be reduced to:

$$\underset{\boldsymbol{d}_r,\boldsymbol{d}_w}{\arg\max} \frac{\mathbb{E}\left\{\boldsymbol{d}_r^T\boldsymbol{r}_c(\boldsymbol{w}_{att}-\boldsymbol{w}_{unt})^T\boldsymbol{d}_w\right\}}{\sqrt{\boldsymbol{d}_r^T\boldsymbol{\Sigma}_{r_cr_c}\boldsymbol{d}_r}\sqrt{\boldsymbol{d}_w^T\boldsymbol{\Sigma}_{ww}\boldsymbol{d}_w}}. \tag{3.13}$$

Provided both stimuli have been normalized, the assumption of equal covariance for the stimuli windows is valid considering that they represent the slow moving temporal amplitude envelopes of speech. That is, relatively close lags will be more correlated with each other than further lags, resulting in a covariance matrix with higher weights around the diagonal. Slight differences in roll-off structure may occur when, for example, relating a fast talker to a slow talker, in which the roll-off around the diagonal would be much sharper for the fast talker; however, this was not an issue for the stimuli used in this study.

The assumption of equal covariance for the stimuli is important as it allows for the LDA based formulation of equation 3.12 to now be posed within the standard framework of CCA. Similar to the stimuli-difference method of SDMMSE, defining $\boldsymbol{w}_{au} = \boldsymbol{w}_{att} - \boldsymbol{w}_{unt}$ allows us to represent this LDA based criterion as:

$$\begin{aligned} \boldsymbol{d}_{r-SDCCA}, \boldsymbol{d}_{w-SDCCA} = \underset{\boldsymbol{d}_r, \boldsymbol{d}_w}{\arg\max} \quad & \sqrt{2}\mathbb{E}\left\{\boldsymbol{d}_r^T \boldsymbol{r}_c \boldsymbol{w}_{au}^T \boldsymbol{d}_w\right\} \\ \text{subject to } & \mathbb{E}\{\boldsymbol{d}_r^T \boldsymbol{r}_c \boldsymbol{r}_c^T \boldsymbol{d}_r\} = 1 \\ & \mathbb{E}\{\boldsymbol{d}_w^T \boldsymbol{w}_{au} \boldsymbol{w}_{au}^T \boldsymbol{d}_w\} = 1. \end{aligned} \tag{3.14}$$

We refer to this multimodal projection technique as stimuli-difference CCA (SDCCA). The gain of $\sqrt{2}$ is introduced due to expanding the stimulus projection constraint, where it can be found that $\mathbb{E}\{\boldsymbol{d}_w^T \boldsymbol{w}_{att} \boldsymbol{w}_{att}^T \boldsymbol{d}_w\} = \mathbb{E}\{\boldsymbol{d}_w^T \boldsymbol{w}_{unt} \boldsymbol{w}_{unt}^T \boldsymbol{d}_w\} = 1/2$, by making another reasonable assumption that both stimuli windows are uncorrelated, i.e. $\mathbb{E}\{\boldsymbol{w}_{att} \boldsymbol{w}_{unt}^T\} = \boldsymbol{0}$.

Similar to SDMMSE, this optimization criterion aims to find projections which produce correlations orthogonal to the GRAAD detection threshold line. Therefore, projections trained to detect an attended speech source can be used to detect an unattended speech source simply by negating the stimulus window projection (i.e. $\boldsymbol{d}_{w-SDCCA}^{att} = -\boldsymbol{d}_{w-SDCCA}^{unt}$ and $\boldsymbol{d}_{r-SDCCA}^{att} = \boldsymbol{d}_{r-SDCCA}^{unt}$). Also similar to SDMMSE, this technique can easily generalize to multiple unattended talkers simply by summing across all normalized unattended stimuli when training the projections (i.e. $\boldsymbol{w}_{au} = \boldsymbol{w}_{att} - \sum_{j=1}^{N_s-1} \boldsymbol{w}_{unt}^{(j)}$, where j denotes the unattended talker index). As stated for SDMMSE, we do not suggest the brain encodes multiple unattended stimuli as this linear sum for large N_s.

Chapter 4

RESULTS

4.1 Experimental Setup

The data used for evaluation in this thesis was generously provided by Prof. Edmund C. Lalor of Trinity Centre for Bioengineering and Trinity College Institute of Neuroscience, Trinity College Dublin. These data have been published previously using a different analysis approach in [28, 30], as well as in [27] where the SRAAD technique based on MMSE estimation was first introduced. The participants, stimuli and procedures, as well as data and preprocessing used in this thesis follow that of [28], which is stated with minimal modifications below.

4.1.1 Participants

34 human subjects took part (mean ± s.d. age, 27.3 ± 3.2 years; 28 male; 7 left-handed). The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics Committee of the School of Psychology at Trinity College Dublin approved the experimental procedures and each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder. (Note however, that [27, 30] report 40 subjects, as opposed to the 34 used here.)

4.1.2 Stimuli and procedures

Subjects undertook 30 trials, each of ~60 seconds in length, where they were presented with 2 classic works of fiction: one to the left ear and the other to the right ear. Each story was read by a different male speaker and, for both stories, each trial began where the story ended on the previous trial. Each subject attended to the story in either their left or right

ear throughout all 30 trials (17 subjects to the left and 17 subjects to the right). After each trial, subjects were required to answer between 4 and 6 multiple-choice questions on both stories. Each question had 4 possible answers. (See [27] for further information). As reported previously [27, 30], the behavioral results clearly showed that subjects were compliant in the task. On average, subjects correctly answered 80.4 $\pm$ 7.3% of the attended questions and 27.1 $\pm$ 7.0% of the unattended, which was not statistically greater than chance ($p = 0.77$).

4.1.3 Data acquisition and preprocessing

EEG data were recorded using $C = 128$ electrode positions. The data were filtered over the range 0-134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. Data were then filtered offline between 2 and 8 Hz in both a forward and backward direction to remove phase distortion and re-referenced to the average of all scalp electrode channels. In order to decrease the processing time required, all EEG data were then downsampled by a factor of 8 to give an equivalent sampling rate of 64 Hz.

The amplitude envelope of the speech stimuli were obtained using a Hilbert transform where it was then low-pass filtered to 8 Hz and decimated to a sampling rate of 64 Hz. EEG responses were evaluated from 0 to 250 ms poststimulus, resulting in $L = 17$. All stimuli and responses were demeaned and properly normalized according to the decoder training technique being evaluated.

4.1.4 Decoder training

Decoders were trained using all but one trial, where testing was evaluated on the remaining trial using a leave-one-out cross-validation approach. Similar to [7], decoders were trained by concatenating all training trials together so better behaved estimations of covariance matrices could be obtained (as opposed to averaging decoders together learned on a trial-by-trial basis). This concatenation of trials is acceptable due to each trial consisting of unique EEG responses and stimuli. Covariance inversion using PCA or regularization for decoder training was not performed on these data.

4.2 Detection and Decoding Accuracy

Correlation coefficients were obtained by measuring the similarity of the decoded neural response with the true speech envelope (or projected speech envelope window for CCA and SDCCA) over the entire test trial duration (~60 seconds). For each method, decoders were trained for the detection of both attended and unattended stimuli, resulting in a total of 12,240 decoders (2 stimuli $\times$ 30 trials $\times$ 34 subjects $\times$ 6 decoders, where the 6 decoders represent 1 for both MMSE and SDMMSE methods and 2 for both CCA and SDCCA methods).

4.2.1 Detection accuracy

We define detection accuracy as the number of trials correctly classified according to equation 2.7 for MMSE and SDMMSE (e.g., $\frac{1}{30}\sum_{k=1}^{30} I_{att}[k]$), or equation 3.8 for CCA and SDCCA (e.g., $\frac{1}{30}\sum_{k=1}^{30} I_{att}^{sub}[k]$). The accuracy at which classification performance is deemed significantly greater than chance is 63.33% based on a binomial test at the $\alpha = 5\%$ significance level. The average detection accuracy across subjects is shown in table 4.1, where individual subject performance is shown in figure 4.1.

Detection Accuracy		
method	s_{att} detection	s_{unt} detection
MMSE (baseline)	77.59 ± 13.46%	69.64 ± 10.15%
SDMMSE	83.19 ± 10.05%	83.19 ± 10.05%
CCA	80.72 ± 15.18%	71.32 ± 14.83%
SDCCA	83.99 ± 12.55%	83.99 ± 12.55%

Table 4.1: Average detection accuracy ± s.d. across all subjects for detecting (**left**) an attended speech stimulus and (**right**) an unattended speech stimulus.

The baseline of MMSE estimation was comparable to that of previously reported studies

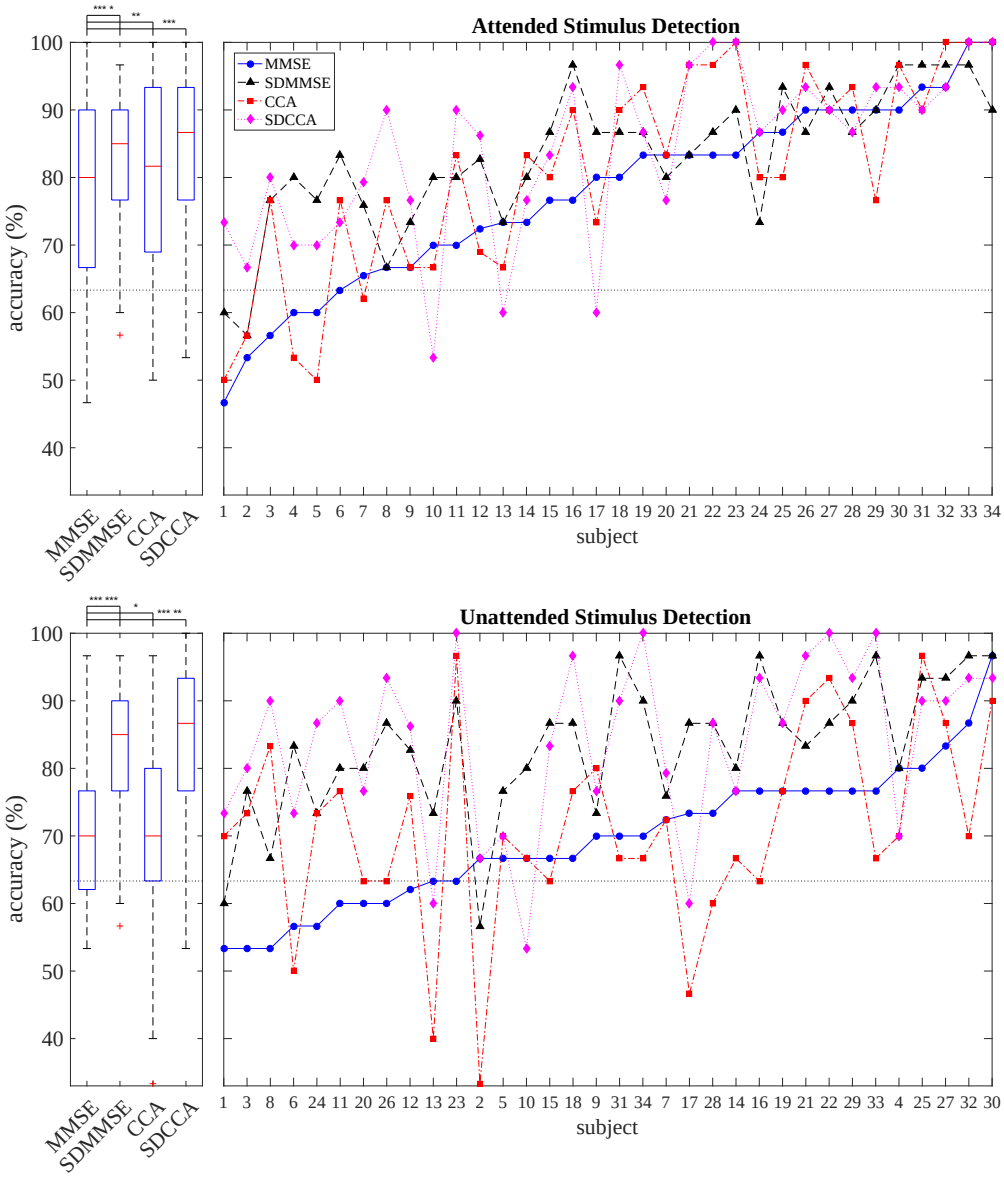

Figure 4.1: Subject specific detection accuracy for each decoding method. The dotted line indicates the classification performance level at which detection-accuracy is significantly greater than chance (63.33%) based on a binomial test at the 5% significance level. **Top**: detection accuracy of an attended speech stimulus. **Bottom**: detection accuracy of an unattended speech stimulus. $p > 5 \times 10^{-2}$ (*), $p < 5 \times 10^{-2}$ (**), $p < 5 \times 10^{-3}$ (***), $p < 5 \times 10^{-4}$ (*** *), $p < 5 \times 10^{-5}$ (*** **), $p < 5 \times 10^{-6}$ (*** ***).

[27, 28], where 2 of the 34 subjects had 100% detection accuracy of an attended stimulus and 29 were significantly above chance. For detection of an unattended stimulus, no subjects achieved 100% detection accuracy, and 25 subjects were significantly above chance. Varying results compared to previously reported accuracies can be attributed to differences in decoder training and preprocessing of the stimuli and EEG.

The proposed LDA based stimuli-difference decoding scheme of SDMMSE resulted in no subjects achieving 100% detection rates for either stimulus. However, the number of subjects identified as significantly above chance increased to 32 for detection of both attended and unattended stimuli when compared to the baseline. SDMMSE resulted in a significant increase in overall detection accuracy for an attended stimulus ($p = 2.95 \times 10^{-4}$; right-tailed Wilcoxon Signed Rank Test), as well as detection of an unattended stimulus ($p = 8.96 \times 10^{-7}$) when compared to the baseline.

The low-rank decoding scheme of GRAAD which utilizes CCA for maximizing the correlation of the projected spatial neural response and a projected stimulus envelope window (without accounting for both attentional stimuli) resulted in 4 subjects achieving perfect detection accuracy of an attended stimulus and none for the unattended stimulus. There were 29 subjects significantly above chance for detection of both attended and unattended stimuli. CCA resulted in a significant increase in overall detection accuracy for an attended stimulus ($p = 2.65 \times 10^{-2}$), however, did not yield a significant increase for an unattended stimulus ($p = 0.22$) when compared to the baseline of MMSE.

The proposed low-rank stimuli-difference decoding scheme of SDCCA resulted in 4 subjects achieving perfect detection accuracy, as well as 29 subjects who were significantly above chance, for detection of both attended and unattended stimuli. A significant increase in overall detection accuracy was achieved for detecting an attended stimulus ($p = 1.70 \times 10^{-3}$), as well as detecting an unattended stimulus ($p = 1.12 \times 10^{-5}$) when compared to the baseline of MMSE.

It should be noted that detection accuracies for both stimuli-difference schemes (SDMMSE and SDCCA) resulted in exactly the same detection accuracies for both attended

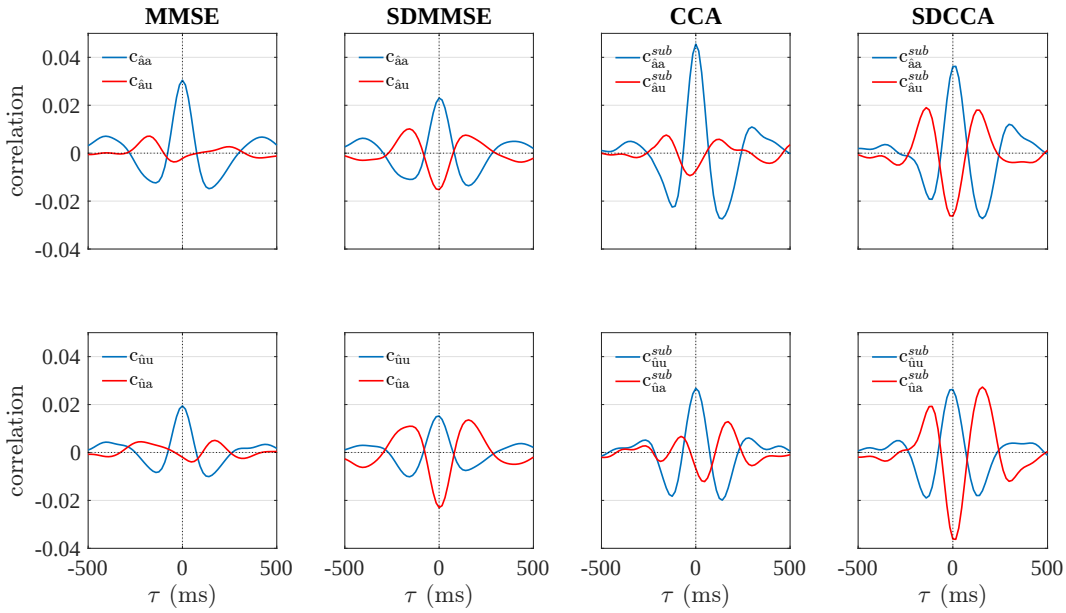

Figure 4.2: Decoding accuracy for each method discussed. Each plot details the average sample cross-correlation of the decoded neural response with the true speech envelope (or projected window of the speech envelope for CCA and SDCCA). Correlations are averaged across all subjects and trials. Sample cross-correlation is defined as $c_{\hat{x},y}(\tau) = \text{corr}\{\hat{s}_x(t), s_y(t-\tau)\}$ or $c^{sub}_{\hat{x},y}(\tau) = \text{corr}\{\hat{\psi}_x(t), \psi_y(t-\tau)\}$, where $\text{corr}\{\cdot\}$ is the sample Pearson correlation coefficient of equation 2.6. **Top row**: decoding accuracy for predicting an attended speech envelope. **Bottom row**: decoding accuracy for predicting an unattended speech envelope.

and unattended stimuli. This is due to the "one-fits-all" decoders that are learned, where selection of attentional detection is performed by decoder negation (effectively negating each of the correlation coefficients obtained depending on attentional decoder selection).

4.2.2 Decoding accuracy

We define decoding accuracy by how well the decoded neural response correlates with the true stimulus[1] (or the projected stimulus window for CCA and SDCCA). The decoding accuracy for each method, averaged across all subjects and trials, is shown in figure 4.2. This figure

[1]Correlation coefficients used for discussing decoder accuracy are defined at $\tau = 0$ when referencing the cross-correlation comparison of figure 4.2.

shows how each decoding method encodes (or omits) information relating to both stimuli within the prediction. For example, when comparing MMSE to SDMMSE for the detection of an unattended stimulus (bottom left two plots of figure 4.2), there is a clear decrease with SDMMSE in average correlation of the prediction with a true unattended stimulus. This decrease is due to optimization criterion which instead focuses on maximizing the distance between $c_{\hat{u},u}$ and $c_{\hat{u},a}$, a distance which is clearly more separated when compared to MMSE. In fact, each of the stimuli-difference techniques showed improvements in correlation separation when compared to MMSE, for both attended and unattended stimulus detection. This is important as it directly affects the detection statistics for both the SRAAD and GRAAD frameworks. The average correlation coefficients obtained across subjects, as well as the average difference between the detection statistic correlations, are shown in table 4.2. Figure 4.3 provides an alternative look at these comparisons by showing the mean correlation coefficient obtained for each subject and how they relate to the detection statistics for both the SRAAD and GRAAD frameworks.

Decoding Accuracy		
method	average correlation (separation) for $\hat{s} = \hat{s}_{att}$ or $\widehat{\psi} = \widehat{\psi}_{att}$	average correlation (separation) for $\hat{s} = \hat{s}_{unt}$ or $\widehat{\psi} = \widehat{\psi}_{unt}$
MMSE (baseline)	0.030 ± 0.030 (0.033 ± 0.041)	0.019 ± 0.028 (0.021 ± 0.039)
SDMMSE	0.023 ± 0.029 (0.038 ± 0.040)	0.015 ± 0.028 (0.038 ± 0.040)
CCA	0.046 ± 0.044 (0.053 ± 0.061)	0.027 ± 0.042 (0.033 ± 0.060)
SDCCA	0.036 ± 0.043 (0.062 ± 0.062)	0.026 ± 0.040 (0.062 ± 0.062)

Table 4.2: Average decoding accuracy ± s.d. across all trials and subjects. (·) denotes the difference between the correlation of the predicted stimulus with the true stimulus and the correlation of the prediction with the opposite stimulus (see text).

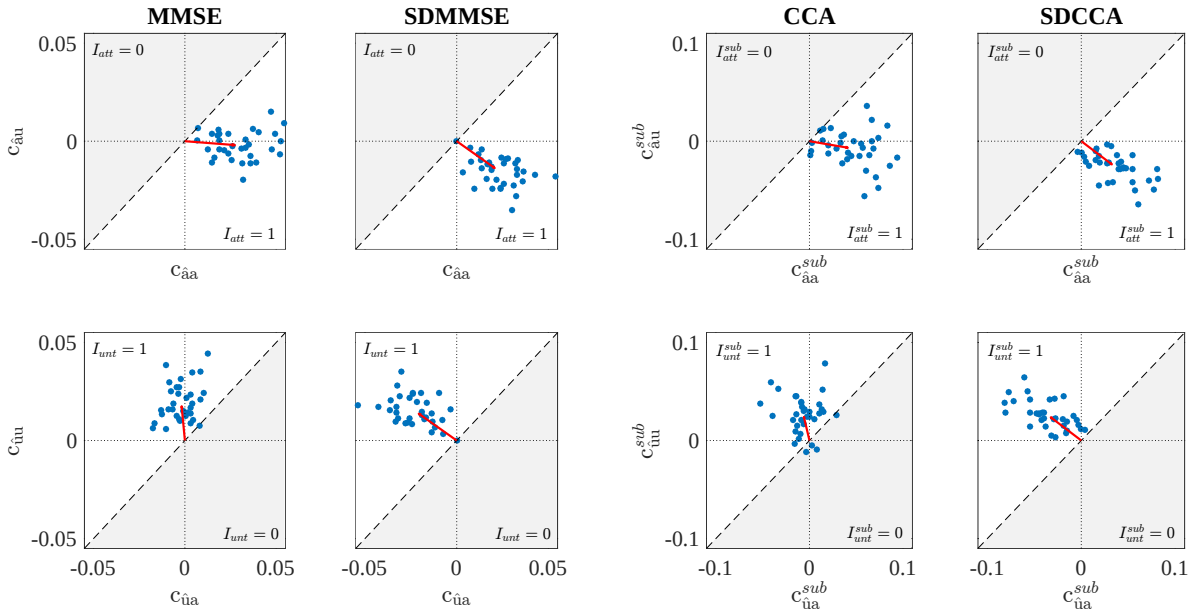

Figure 4.3: The mean correlation coefficients obtained for each subject. The red line indicates direction and magnitude of the trend based on mean correlation over all subjects and trials. **Top row**: correlation coefficients obtained when predicting an attended speech envelope. **Bottom row**: correlation coefficients obtained when predicting an unattended speech envelope.

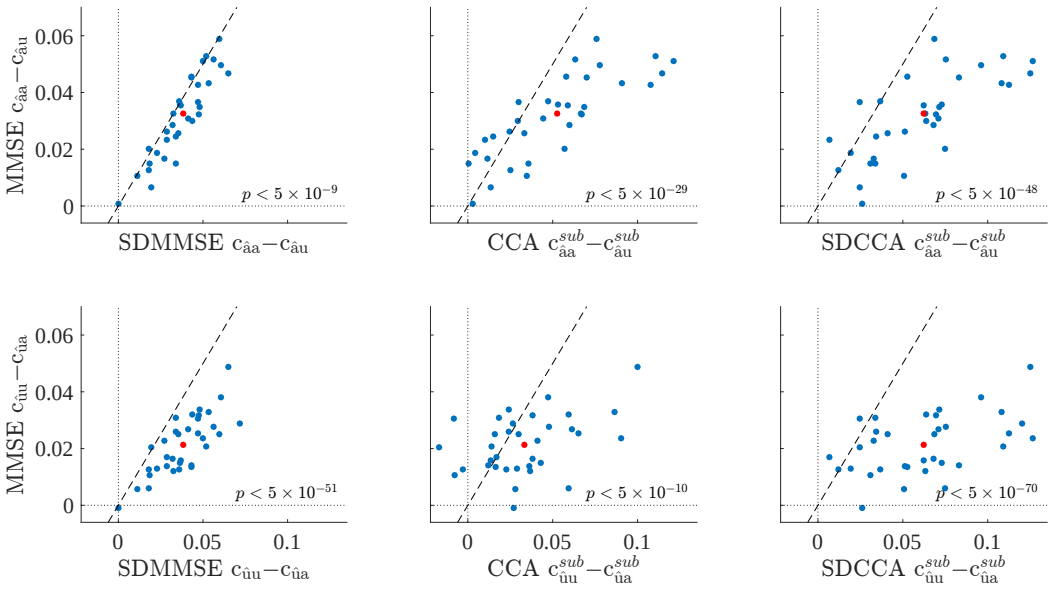

Figure 4.4: Comparison of each proposed decoding method against traditional MMSE estimation. Blue dots denote the mean difference between detection statistic correlations obtained for each subject. The red dot indicates mean correlation difference over all subjects and trials. **Top row**: correlation difference obtained when predicting an attended speech envelope. **Bottom row**: correlation difference obtained when predicting an unattended speech envelope.

Because the expected distance between correlation coefficients is a central part for both SRAAD and GRAAD detection statistics, we compared this metric against the baseline of MMSE for each of the proposed methods. The results are shown in figure 4.4. There were significant increases in correlation separation when predicting an attended stimulus using SDMMSE, CCA, and SDCCA ($p < 5 \times 10^{-9}$, $p < 5 \times 10^{-29}$, and $p < 5 \times 10^{-48}$, respectively; Wilcoxon Signed Rank Test), as well as for predicting an unattended stimulus ($p = 5 \times 10^{-51}$, $p < 5 \times 10^{-10}$, and $p < 5 \times 10^{-70}$, respectively).

Chapter 5

CONCLUSION

In this thesis, two frameworks for auditory attention detection (AAD) were examined. In chapter 2, the method of auditory attention detection via stimulus reconstruction (SRAAD) was discussed. Here it was introduced how the method of stimulus reconstruction had previously been applied to AAD, where neural decoding functions were regressed toward the speech envelope of the desired attentional speech source (i.e. an attended or unattended speech source) during training using minimum-mean square error (MMSE) criterion. It was suggested that this type of reconstruction discards useful information relating to the opposite attentional stimulus (attended vs. unattended) which could otherwise be beneficial to auditory attention detection. It was then proposed how the neural encoding properties to both attended and unattended stimuli could be jointly accounted for by adapting concepts inherent to linear discriminant analysis (LDA). Here, the optimization criterion was proposed to learn decoders which produce correlations orthogonal to the detection threshold line for SRAAD, effectively maximizing the attentional class separation between correlation coefficients used in the detection statistic to improve accuracy. It was shown that this concept reduces to a modified version of MMSE estimation, where regression for the neural decoder is performed on the difference of normalized attentional stimuli. We referred to this as stimuli-difference MMSE (SDMMSE) estimation. Experimental validation suggested significant improvement in accuracy for the detection of both attended and unattended speech stimuli when compared to traditional MMSE stimulus reconstruction.

The second type of AAD framework was proposed in chapter 3. Here it was introduced how the spatiotemporal relationship between the neural response and speech envelope could be split between modalities, where auditory attention detection could be performed

by predicting the subspace of a projected stimulus window using only the spatial component of the neural response. This resulted in a low-rank alternative to standard SRAAD techniques which we referred to as generalized reconstruction for auditory attention detection (GRAAD). It was then derived how canonical correlation analysis (CCA) can be used for learning the optimal decoder for the spatial neural response, as well as the optimal subspace projection needed for the stimulus window. Experimental analysis suggested this approach is comparable to that of standard MMSE for the detection of an attended stimulus, however, had reduced performance when detecting an unattended stimulus. We attributed this reduction in accuracy to the reduced number of neural response features used within the GRAAD framework. To counteract this problem, it was then shown how the LDA inspired method of SDMMSE could be adapted into this multimodal projection framework. Here, optimization criterion was proposed with the goal of maximizing the attentional class separation between the correlation coefficients used in the GRAAD detection statistic. We found that this criterion can be posed within traditional CCA criterion, where stimulus projections are learned on the difference between temporal lag windows of both attentional stimuli. We referred to this method as stimuli-difference CCA (SDCCA). This yielded the highest increase in average detection accuracy across subjects compared to the baseline of MMSE for the detection of both attended and unattended stimuli.

We have shown that accounting for the expected neural encoding properties for both attended and unattended stimuli within the stimulus reconstruction technique results in improved reconstruction statistics for the detection of attentional stimuli. We have also shown that high-dimensional spatiotemporal neural responses are not necessary when using the GRAAD framework to achieve detection accuracy comparable to traditional stimulus reconstruction approaches. A promising area for future research is to consider a more probabilistic approach to stimulus reconstruction. Currently, methods based on linear regression (as with the methods proposed in this thesis) are able to learn the expected long-term encoding properties of the neural population so that reasonable reconstruction accuracy can be achieved over extended periods (~60 seconds). However, this accuracy tends to break

down when evaluating over shorter time windows (shown in [27] and [28] to decrease linearly with evaluation window length). By developing reconstruction techniques which attempt to learn short-term dynamic interaction between the stimuli and neural responses, where more intricate data-driven concepts of machine learning are used (e.g. dictionary learning techniques for probabilistic decoder training), we hypothesize for the future that more detailed reconstruction can be achieved to allow for better AAD performance.

BIBLIOGRAPHY

[1] Steven J. Aiken and Terence W. Picton. Human cortical responses to the speech envelope. *Ear and Hearing*, 29(2):139–157, Apr. 2008.

[2] Sahar Akram, Alessandro Presacco, Jonathan Z Simon, Shihab A Shamma, and Behtash Babadi. Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. *NeuroImage*, 124:906–917, 2016.

[3] Sahar Akram, Jonathan Z Simon, Shihab A Shamma, and Behtash Babadi. A state-space model for decoding auditory attentional modulation from meg in a competing-speaker environment. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 460–468. Curran Associates, Inc., 2014.

[4] Ali Aroudi, Bojana Mirkovic, Maarten De Vos, and Simon Doclo. Auditory attention decoding with EEG recordings using noisy acoustic reference signals. In *Proc. of the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016. IEEE.

[5] Francis R Bach and Michael I Jordan. A probabilistic interpretation of canonical correlation analysis. 2005.

[6] William Bialek, Fred Rieke, Rob R. de Ruyter Van Steveninck, and David Warland. Reading a neural code. *Science*, 252(5014):1854–1857, June 1991.

[7] Wouter Biesmans, Jonas Vanthornhout, Jan Wouters, Marc Moonen, Tom Francart, and Alexander Bertrand. Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario. In *Proc. of the 37th Annual Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5155–5158, Milan, Italy, Aug. 2015. IEEE.

[8] Åke Björck and Gene H Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of computation*, 27(123):579–594, 1973.

[9] E Colin Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979, 1953.

[10] Neetha Das, Simon Van Eyndhoven, Tom Francart, and Alexander Bertrand. Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses (preprint).

[11] Nai Ding and Jonathan Z. Simon. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. of the National Academy of Sciences of the United States of America (PNAS)*, 109(29):11854–11859, June 2012.

[12] Nai Ding and Jonathan Z. Simon. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 107(1):78–89, Jan. 2012.

[13] Bradley Ekin, Les Atlas, Majid Mirbagheri, and Adrian KC Lee. An alternative approach for auditory attention tracking using single-trial EEG. In *Proc. of the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016. IEEE.

[14] Gene Howard Golub. Matrix decompositions and statistical calculations. 1969.

[15] Elana M. Zion Golumbic, Nai Ding, Stephan Bickel, Peter Lakatos, Catherine A. Schevon, Guy M. McKhann, Robert R. Goodman, Ronald Emerson, Ashesh D. Mehta, Jonathan Z. Simon, David Poeppel, and Charles E. Schroeder. Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party. *Neuron*, 77(5):980–991, Mar. 2013.

[16] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[17] Jon R Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

[18] Edmund C. Lalor and John J. Foxe. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, 31(1):189–193, Dec. 2010.

[19] Timo Lauteslager, James A. O'Sullivan, Richard B. Reilly, and Edmund C. Lalor. Decoding of attentional selection in a cocktail party environment from single-trial EEG is robust to task. In *Proc. of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1318–1321, Chicago, IL, USA, Aug. 2014.

[20] David Looney, Cheolsoo Park, Yili Xia, Preben Kidmose, Michael Ungstrup, and Danilo P Mandic. Towards estimating selective auditory attention from EEG using

a novel time-frequency-synchronisation framework. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–5. IEEE, 2010.

[21] Thomas Lunner. About cognitive outcome measures at ecological signal-to-noise ratios and cognitive-driven hearing aid signal processing. *American journal of audiology*, 24(2):121–123, 2015.

[22] Huan Luo and David Poeppel. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6):1001–1010, 2007.

[23] Nima Mesgarani and Edward F. Chang. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236, May 2012.

[24] Nima Mesgarani, Stephen V David, Jonathan B Fritz, and Shihab A Shamma. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of neurophysiology*, 102(6):3329–3339, 2009.

[25] Majid Mirbagheri, Bradley Ekin, Les Atlas, and Adrian KC Lee. Flexible tracking of auditory attention. In *Proc. of the Sixteenth Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Dresden, Germany, Sept. 2015.

[26] Bojana Mirkovic, Stefan Debener, Manuela Jaeger, and Maarten De Vos. Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of Neural Engineering*, 12(4):046007, June 2015.

[27] James A. O'Sullivan, Alan J. Power, Nima Mesgarani, Siddharth Rajaram, John J. Foxe, Barbara G. Shinn-Cunningham, Malcolm Slaney, Shihab A. Shamma, and Edmund C. Lalor. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, page bht355, 2014.

[28] James A O'Sullivan, Richard B Reilly, and Edmund C Lalor. Improved decoding of attentional selection in a cocktail party environment with EEG via automatic selection of relevant independent components. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 5740–5743. IEEE, 2015.

[29] Brian N. Pasley, Stephen V. David, Nima Mesgarani, Adeen Flinker, Shihab A. Shamma, Nathan E. Crone, Robert T. Knight, and Edward F. Chang. Reconstructing speech from human auditory cortex. *PLoS-Biology*, 10(1):175, Jan. 2012.

[30] Alan J Power, John J Foxe, Emma-Jane Forde, Richard B Reilly, and Edmund C Lalor. At what time is the cocktail party? a late locus of selective attention to natural speech. *European Journal of Neuroscience*, 35(9):1497–1503, 2012.

[31] Alexandro D Ramirez, Yashar Ahmadian, Joseph Schumacher, David Schneider, Sarah MN Woolley, and Liam Paninski. Incorporating naturalistic correlation structure improves spectrogram reconstruction from neuronal activity in the songbird auditory midbrain. *The Journal of Neuroscience*, 31(10):3828–3842, 2011.

[32] F Rieke, DA Bodnar, and W Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London B: Biological Sciences*, 262(1365):259–265, 1995.

[33] Sarah A Simpson and Martin Cooke. Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America*, 118(5):2775–2778, 2005.

[34] Garrett B. Stanley, Fei F. Li, and Yang Dan. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *The Journal of Neuroscience*, 19(18):8036–8042, Sept. 1999.

[35] Simon Van Eyndhoven, Tom Francart, and Alexander Bertrand. EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *arXiv preprint arXiv:1602.05702*, 2016.

[36] David K. Warland, Pamela Reinagel, and Markus Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78(5):2336–2350, Nov. 1997.

VITA

Bradley Ekin was raised in Michigan where he gained an interest in electrical engineering at an early age working on projects at his father's machine shop. Gaining a deeper interest in signal processing and embedded systems in his undergraduate career at Lake Superior State University, today he pursues research topics in machine learning, optimization, acoustic signal processing, as well as a recent found interest in computational neuroscience. Brad is continuing on as a PhD student in the Department of Electrical Engineering at the University of Washington.