Characterising the Human Auditory System using a Linear Least-Squares System Identification Approach

Denis P. Drennan, BEng, MEng

Under the supervision of

Dr. Edmund C. Lalor

A dissertation submitted to the

University of Dublin, Trinity College

in fulfilment of the requirements for the degree of

Doctor of Philosophy

16th August 2019

Department of Electronic and Electrical Engineering

University of Dublin, Trinity College

Declaration

I, Denis Drennan, confirm that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed,

Denis Drennan

Denis P. Drennan

16th August 2019

Summary

Disabling hearing loss affects many millions of people around the world. Early identification and suitable interventions, e.g., the provision of hearing aids, cochlear implants, etc., can help but are limited by the methods currently used to assess hearing function. Hearing function can be assessed in one of two ways: using either subjective or objective methods. Subjective methods rely on the behavioural response to sound, while objective methods rely on the physiological response to sound. While subjective methods such as pure tone audiometry (PTA) have played an important role in hearing assessment for decades, they are generally limited to simple detection tasks, e.g. detecting pure tones in quiet, and provide little indication as to the source of any deficit. While some objective measures such as the auditory evoked potential (AEP)—a transient electrophysiological response to sound—can provide more detailed neurophysiological information, they are often restricted by the need to use simple discrete stimuli such as clicks or tone-bursts, which are arguably not that representative of everyday sounds.

In recent years, there has been growing interest in the use of modelling approaches to study and assess the human auditory system. One approach that has proven particularly useful is temporal response function (TRF) estimation. With TRF estimation, the assumption is that the output neural data, consists of the convolution of some input stimulus feature with an unknown system response, plus noise. Given the known stimulus feature and the recorded neural response, the intermediary system response (TRF), can be derived using TRF estimation. One of the main advantages of this approach is that it provides similar responses to the AEP, while permitting a wider variety of stimuli to be used. Much of the work that has been done using this approach has been focused on the study and assessment of high-level (cognitive) auditory processing. The overall aim of

this thesis is to develop and appraise new methodological approaches that facilitate the use of TRF estimation in the study and assessment of low-level (sensory) auditory processing.

In Chapter 4, two approaches for indexing low-level processing along the auditory pathway are introduced over two experiments: Experiment 1 and 2. Experiment 1 is an initial exploratory attempt to derive responses along the auditory pathway to click trains, i.e., sequences of click stimuli—the classic stimuli of auditory research. Experiment 2 is a more thorough attempt to derive responses along the auditory pathway to amplitude modulated (AM) broadband noise (BBN), using a novel efficient TRF estimation approach. Considerations of stimulus type, stimulus representation, i.e., what stimulus feature to use and how to represent it in the analysis, and computational efficiency are discussed, and the neural underpinnings of the derived responses investigated through comparisons with their canonical counterparts, i.e., AEPs, elicited using chirp trains.

In Chapter 5, a novel TRF estimation approach for objectively determining hearing thresholds, i.e., the lowest levels at which certain sounds can be heard in quiet, using multiplexed, i.e., multiple, mixed, AM tones (AMTs) is presented. Considerations of stimulus type, stimulus representation, and modelling approach are discussed, and the performance of this approach evaluated through comparisons with thresholds recovered using PTA, in both normal hearing and hearing loss populations.

In Chapter 6, several novel stimulus representations are presented with the view of enhancing response derivation using TRF estimation. The importance and benefits of taking certain neurophysiological properties of the human auditory system into account when designing stimulus representations are discussed and then quantified through comparisons with other models derived using more standard stimulus representations.

While the focus of this thesis is on low-level auditory processing, any high-level processing necessarily involves some degree of low-level processing as well. Therefore, it is hoped that the work presented in this thesis will help to further the study and assessment of both low- and high-level processing using TRF estimation. As such, it represents an opportunity to move the field closer towards more diagnostically useful objective measures of hearing function along the auditory pathway that can be implemented using a wider variety of stimuli.

Acknowledgements

This thesis would not have been possible without the support of several people. Therefore, I would like to thank:

Ed, for your constant support and guidance throughout the PhD. I have learned a lot from you over the years. Thank you for making it both interesting and fun.

Richard, for all your kind help and advice.

Mick, for getting me set up in the beginning, and for your time and patience even though you were so busy completing your own PhD.

Kevin, for all your help, particularly with MAMTA. It is very much appreciated.

Other past and current members of the Lalor Lab: Ger, Jim, Gio, Andy, Nate, Aaron, Adam, Emily, Michael, Aisling, Joyce, Kathryn, Shyantony, and Lauren. Thank you for all your helpful feedback and advice, and for making the Lalor Lab such a nice place to work.

Ter, for all the consultation and comic relief. This would have been a very different experience without you. Thanks for everything.

Other past and current members of the Reilly Lab: Céline, Isabelle, Martin, Alejandro, Brendan, Niamh, Saskia, Ciara, Surbhi, Shruti, Clodagh, and Eugene. Thank you for making these last few years so memorable.

Ross, for getting me started on this road many years ago, and for your support throughout. I have learnt a lot from you over the years, for which I am very grateful.

Steve and Cormac, the best engineers I know. Thank you for all your help and patience. I have also learnt a lot from you both.

Caroline and Brendan, for all your help with patient recruitment, audiological testing, and for all the good times.

Everyone else in Neuromod Devices for making it a wonderful place to work over the last few years.

My parents Thos and Angela, for your constant love and support, and for instilling within me a strong work ethic, through your own example.

My brother Tom for your friendship and advice, and for feeding my intellectual curiosity outside of the PhD.

Marie, for your love, support, and patience throughout what was a challenging few years at the beginning of our relationship. I will always be grateful.

All the subjects who generously volunteered their time to participate in these studies.

This research was kindly supported by the Irish Research Council (IRC) and Science Foundation Ireland (SFI).

DENIS P. DRENNAN

Trinity College Dublin

August 2019

Publications Arising from this Thesis

Journal Articles

- **Drennan DP**, Lalor EC (2019). Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. eNeuro, ENEURO.0082-19.2019.

Journal Articles in Preparation

- **Drennan DP**, Lalor EC (in prep.). Indexing the Human Auditory Processing Hierarchy: Considerations of Stimulus Type, Stimulus Representation, and Computational Efficiency.

Conference Poster Presentations

- **Drennan DP**, Lalor EC. Improved Modelling of Auditory Responses to Amplitude Modulation using Multivariate Representations of Amplitude. *Auditory EEG Signal Processing 2018*, Leuven, Belgium.
- **Drennan DP**, Lalor EC. A System Identification Approach for Rapid Characterization of the Human Auditory System. *Society for Neuroscience 2016*, San Diego, USA.
- **Drennan DP**, Lalor EC. A System Identification Approach for Rapid Characterization of the Human Auditory System. *Advances and Perspectives in Auditory Neuroscience 2016*, San Diego, USA.

- **Drennan DP**, Cross MJ, Di Liberto GM, Lalor EC. A System Identification Approach for Rapid Characterization of the Ascending Human Auditory Hierarchy. *Association for Research in Otolaryngology 2016,* San Diego, USA.

Table of Contents

List of Figures

List of Tables

Glossary of Abbreviations

AB	Amplitude-Binned
ABR	Auditory Brainstem Response
ADJAR	Adjacent Response
AEP	Auditory Evoked Potential
AM	Amplitude Modulated
AMT	Amplitude Modulated Tone
ASSR	Auditory Steady-State Response
BBN	Broadband Noise
cABR	Complex Auditory Brainstem Response
CLAD	Continuous Loop Averaging Deconvolution
dB	Decibel
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
FAB	Frequency- and Amplitude-Binned
FB	Frequency-Binned
fMRI	Functional Magnetic Resonance Imaging
FSFR	Full Sampling Rate Full Range
FSRR	Full Sampling Rate Reduced Range
GFP	Global Field Power
IIR	Infinite Impulse Response

LAEP	Late Auditory Evoked Potential
LS	Level-Specific / Least-Squares
MAMT	Multiplexed Amplitude Modulated Tone
MAMTA	Multiplexed Amplitude Modulated Tone Audiometry
MEG	Magnetoencephalography
MLR	Middle-Latency Response
MLS	Maximum Length Sequence
MSAD	Multiple-Rate Steady-State Deconvolution
peRETSPL	Peak-to-Peak Equivalent Reference Equivalent Threshold SPL
peSPL	Peak-to-Peak Equivalent SPL
PT	Pure-Tone
PTA	Pure-Tone Audiometry
QSD	Q-Sequence Deconvolution
RSA	Randomized Stimulation and Averaging
RSFR	Reduced Sampling Full Range
RSRR	Reduced Sampling Rate Reduced Range
SEM	Standard Error of the Mean
SNR	Signal-to-Noise Ratio
SOA	Stimulus Onset Asynchrony
SPL	Sound Pressure Level
STD	Standard Deviation
TDA	Time-Domain Averaging
TRF	Temporal Response Function
VSFR	Variable Sampling Rate Full Range
WHO	World Health Organization

Chapter 1. Introduction

1.1 Background

According to the World Health Organisation (WHO) 466 million people worldwide have disabling hearing loss, 34 million of whom are children (WHO, 2018). Disabling hearing loss is defined in this context as hearing loss greater than 40 dB in the better hearing ear of adults, and greater than 30 dB in the better hearing ear of children (WHO, 2018). Early identification and suitable interventions, e.g., the provision of hearing aids, cochlear implants, etc., can help, but are limited by the methods currently used to assess hearing function.

Hearing function can be assessed in one of two ways: using either subjective or objective methods. Subjective methods rely on the behavioural response to sound, while objective methods rely on the physiological response to sound. One widely used subjective method is pure-tone audiometry (PTA; Bunch, 1923, 1922). With PTA, a subject is typically presented with short excerpts of a pure-tone (PT) at a fixed level and asked if they can hear it. The level is then adjusted in descending and ascending runs until the lowest level at which they can hear that tone, i.e., their hearing threshold for that tone, has been determined. This process is then repeated for the other frequencies being tested, until a full audiometric profile has been established (BSA, 2011).

While PTA has been instrumental in the assessment of hearing function for decades, it is not without its shortcomings. As the human auditory system comprises a series of processing stages between the ear and cortex, and dysfunction can arise at any of these stages, assessing hearing function based on behavioural responses alone makes it very difficult to determine the source of any deficits. Furthermore, PTA does not evaluate suprathreshold processing, and is not suitable for use with young children or those with a diminished capacity to respond (Downs

et al., 1966). And so, there has been a growing interest for many years in the use of objective measures.

One widely used objective measure is the so-called auditory evoked potential (AEP). This is a transient electrophysiological response to brief sounds, recorded from the auditory system using electrodes placed on the scalp—a technique referred to as electroencephalography (EEG). AEPs show several characteristic voltage fluctuations known as waves at different latencies that can be linked with different processing stages along the auditory pathway. By examining the properties of these waves, one can hope to pinpoint the locus of hearing dysfunction in a quantitative and objective manner (Musiek et al., 1994). Like PTA, the AEP can also be used to determine hearing thresholds (Tyberghein and Forrez, 1971). Unlike PTA, it can be used to evaluate suprathreshold processing (Ruggles et al., 2011) and is suitable for use with young children or those with a diminished capacity to respond.

While these characteristics make the AEP a potentially attractive tool for assessing hearing function, it has traditionally had two major disadvantages. First, response recovery can be relatively slow as the signal-to-noise ratio (SNR) associated with it is poor—as most of the EEG recorded from the scalp is unrelated to the presented sound—and separate sets of stimuli and resulting EEG are required to assess each hierarchical processing stage (latency) of interest. Second, the AEP—like most objective measures—is typically elicited using simple discrete stimuli, like clicks or tone-bursts, but as the human auditory system has evolved to process a broad spectrum of sounds—from simple discrete stimuli such as the crack of a twig, to complex continuous stimuli such as speech—it would seem fitting that the approach used to interrogate this system be able to utilise such a wide variety of sounds.

While some methods have enabled measures to be recovered using more complex stimuli, e.g., the complex auditory brainstem response (cABR; Skoe and Kraus, 2010)—a transient electrophysiological response to short repeated syllables—the quasi-discrete, repeated nature of its elicitation, is unlikely to engage the kinds of high-level (cognitive) auditory processing commensurate with continuous natural speech. While other methods have enabled measures to be recovered using continuous stimuli, e.g., the auditory steady-state response (ASSR; Galambos et al., 1981)—an oscillatory electrophysiological response usually to regularly amplitude modulated (AM) sounds—this comes at the cost of any temporal resolution in the response, i.e., it is typically analysed in the frequency domain.

In recent years, several methods that facilitate the use of a wider variety of stimuli have been proposed. One method that has proven particularly useful is temporal response function

(TRF) estimation (Lalor et al., 2009). TRF estimation is a modelling approach, that tries to fit models that describe the relationship between a sensory input and neural response, i.e., it uses features of the observed inputs, e.g., auditory stimuli, and outputs, e.g., EEG, to mathematically estimate the intermediary, e.g., auditory, system response (TRF). As mentioned, the main advantage of this approach is that it can be used with almost any auditory stimuli while still retaining full temporal resolution in the response. For example, Lalor et al. (2009) showed that using TRF estimation, it is possible to derive low-level, i.e., sensory, auditory responses to quasi-periodic tone-bursts, AM tones (AMTs), and AM broadband noise (BBN), while Lalor and Foxe (2010) showed that it is possible to derive low-level auditory responses to continuous natural speech.

However, much of the work that has been done using this approach has been focused on high-level auditory processing, e.g., while studying the so-called cocktail party attention problem, i.e., the problem associated with attending to one speaker in a multi-speaker environment (O'Sullivan et al., 2015), as well as linguistic processing at the level of phonemes (Di Liberto et al., 2015) and semantics (Broderick et al., 2018). That was until recently, however, when Maddox and Lee (2018) modified the TRF estimation approach and showed that it is possible to derive low-level responses to continuous natural speech, from the entire auditory pathway simultaneously, i.e., not just from the auditory cortex as in Lalor and Foxe (2010). That is not to say that these two endeavours are entirely unrelated, however, indeed this finding may facilitate the study of interactions between high and low-level auditory processing, thus bridging the gap between these two bodies of work.

It would seem then that this approach holds a lot of potential for the study and assessment of low-level auditory processing—for which it remains relatively underexplored—and that any advances to that end, may benefit the study and assessment of high-level processing as well. As such, it represents an opportunity to move the field closer towards more diagnostically useful objective measures of hearing function along the auditory pathway.

1.2 Aims

The overall aim of this thesis is to develop and appraise new methodological approaches that facilitate the use of TRF estimation in the study and assessment of low-level auditory processing.

1. To develop a novel TRF estimation approach for efficiently deriving low-level responses from the entire auditory pathway.
2. To develop a novel TRF estimation approach for objectively determining hearing thresholds, and to pilot it in both normal hearing and hearing loss populations.
3. To develop more neurophysiologically-inspired stimulus representations to enhance response derivation using TRF estimation.

1.3 Thesis Outline

In Chapter 2, the anatomy and physiology of low-level human auditory processing is introduced. This is followed by a brief overview of its pathophysiology, i.e., a brief overview of the different types of hearing loss and how they manifest. The basic principles of EEG are then described, along with a discussion of its strengths and weaknesses. Some of the various measures of low-level auditory processing that can be recovered using EEG are outlined, with a focus on the AEP, as it can be analogous to the TRF in some cases. Finally, some of the different methods that can be used to recover AEPs from EEG are described.

In Chapter 3, the TRF estimation approach is introduced. This is followed by a description of its theoretical/mathematical basis, including definitions of relevant terminology, equations, and procedures. Some nomenclature associated with different types of stimulus representations are then defined, followed by a discussion on how to interpret the resulting TRFs. Finally, the relationship between the TRF and AEP is outlined.

In Chapter 4, two approaches for indexing low-level processing along the auditory pathway are introduced over two experiments: Experiment 1 and 2. Experiment 1 is an initial exploratory attempt to derive responses along the auditory pathway to click trains, i.e., sequences of click stimuli—the classic stimuli of auditory research. Experiment 2 is a more thorough attempt to derive responses along the auditory pathway to AM BBN, using a novel efficient TRF estimation approach. Considerations of stimulus type, stimulus representation, i.e., what stimulus feature to use and how to represent it in the analysis, and computational efficiency are discussed, and the neural underpinnings of the derived responses investigated through comparisons with their canonical counterparts, i.e., AEPs, elicited using chirp trains.

In Chapter 5, a novel TRF estimation approach for objectively determining hearing thresholds using multiplexed, i.e., multiple, mixed, AMTs is presented. Considerations of stimulus type, stimulus representation, and modelling approach are discussed, and the

performance of this approach evaluated through comparisons with thresholds recovered using PTA, in both normal hearing and hearing loss populations.

In Chapter 6, several novel stimulus representations are presented with the view of enhancing response derivation using TRF estimation. The importance and benefits of taking certain neurophysiological properties of the human auditory system into account when designing stimulus representations are discussed and then quantified through comparisons with other models derived using more standard stimulus representations.

Chapter 2. Electrophysiology of Human Auditory Processing

In this chapter, the anatomy and physiology of low-level human auditory processing is introduced. This is followed by a brief overview of its pathophysiology. The basic principles of EEG are then described, along with a discussion of its strengths and weaknesses. Next, some of the various measures of low-level auditory processing that can be recovered using EEG are presented, with a focus on the AEP. Finally, some of the different methods that can be used to recover AEPs from EEG are outlined.

2.1 Anatomy and Physiology of Human Auditory Processing

2.1.1 Peripheral Auditory System

The peripheral auditory system can be broken down into three subsections, specifically the outer-, middle-, and inner-ear (Figure 2.1).

The outer-ear consists of a cartilaginous flange called the pinna, which includes a resonant cavity known as the concha, together with a narrow tube called the auditory (ear) canal that ends with a flexible membrane known as the tympanic membrane (ear drum; Figure 2.1A; Pickles, 2008, p.11). Sound waves are collected at the pinna and directed into the concha. Due to the intricate shape of the pinna, these reflections impart spectral cues that help with sound localisation. These sound waves then continue down along the ear canal, where they get amplified by resonances, predominantly in the 2–7 kHz range—within which much of the frequency content of speech can be found. On reaching the end of the ear canal, the sound waves impinge upon the tympanic membrane, causing it to vibrate.

The middle-ear is an air-filled cavity, comprising three small bones known as the ossicles, i.e., the malleus, incus, and stapes, that contact another flexible membrane on the cochlea, known as the oval window (Figure 2.1A). As sound waves impinge upon the tympanic membrane, the resulting vibrations are transmitted mechanically through the ossicles to the oval window. The main purpose of the middle-ear is to act like a transformer, matching the impedance of the ear canal to the much higher impedance of the cochlear fluids. This is predominantly achieved by virtue of the scaling down—in terms of area—between the tympanic membrane and the oval window—or more accurately the stapes footplate on the oval window—which increases the pressure at the latter (Pickles, 2008, p. 15).

The inner-ear is made up of a snail-shaped tubular structure embedded in the temporal bone called the cochlea, and its connections to the auditory nerve (Figure 2.1A). If the cochlea were cut cross-sectionally, one would see that the tube is divided into three fluid-filled chambers, i.e., the scala vestibuli, the scala media, and the scala tympani (Figure 2.1B). The scala media and scala tympani are separated by the basilar membrane, upon which sits the organ of Corti, and over which hangs the tectorial membrane. The organ of Corti contains ~20,000 hair cells, so-called as they have hair-like stereocilia extending from their top (Figure 2.1C; Bear et al., 2007, p. 354). Hair cells can be divided into inner and outer hair cells depending on their location with respect to the rods of Corti, and are innervated by auditory nerve fibres with their cell bodies in the spiral ganglion (Bear et al., 2007, p. 354; Pickles, 2008, p. 73).

As mechanical vibrations are transmitted through the ossicles to the oval window, the fluids inside the cochlea are perturbed. This perturbation causes the basilar membrane to bend near its base, i.e., near the oval window, starting a travelling wave that propagates towards the apex. As the basilar membrane is narrower and stiffer at its base than at its apex, different vibrational frequencies cause maximal displacement at different points along its length, i.e., it is tonotopically organised, with high-frequencies causing maximal displacement closer to the base and vice versa. As the basilar membrane is displaced by this travelling wave, the affixed hair cells in the organ of Corti are moved either in or out with respect the tectorial membrane. This movement causes the stereocilia to bend, depolarising the hair cells, and causing electrical signals to propagate towards the spiral ganglion. Interestingly, even though there are fewer of them, it is predominantly the inner hair cells that are innervated by the auditory nerve. The outer hair cells contain motor proteins that can cause them to push or pull on the basilar membrane to compensate for low-intensity sounds, thus forming a kind of cochlear amplifier

(Bear et al., 2007, p. 354). The cochlea can be thought of as an analogue filter bank, that splits incoming soundwaves into logarithmically-spaced frequency bands.

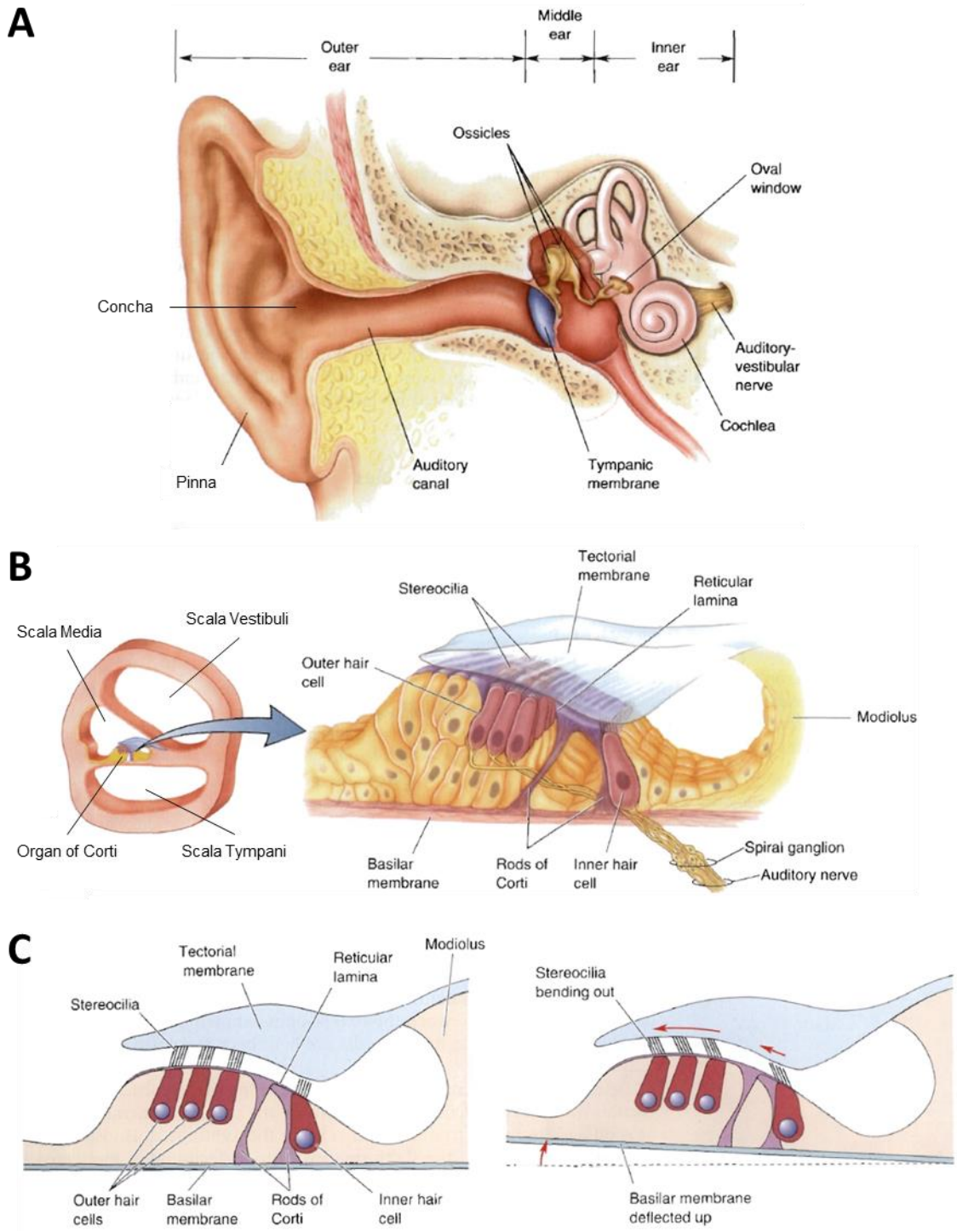

Figure 2.1: Peripheral Auditory System

A *– A diagram of the peripheral auditory system, with the extent of the outer-, middle-, and inner-ear outlined.* ***B*** *– Diagrams of the cross-sectional area of the cochlea, and organ of Corti, respectively.* ***C*** *– Detailed diagrams of the organ of Corti, including hair cells and the effect of basilar membrane deflection. Adapted from Bear et al. (2007).*

2.1.2 Central Auditory System

Electrical signals leaving the cochlea travel to the cochlear nucleus via the auditory nerve (Figure 2.2). The cochlear nucleus then projects to higher nuclei through two main streams: the ventral and dorsal streams. Processes in the ventral stream are predominantly concerned with sound localisation, while processes in the dorsal stream are predominantly concerned with sound identification (Pickles, 2008, p. 155). The ventral (sound localisation) stream runs to the superior olivary nuclei on both sides, and up through the lateral lemniscus to the inferior colliculus, while the dorsal (sound identification) stream runs primarily through the lateral lemniscus to the inferior colliculus on the opposite side (Pickles, 2008, p. 170). The inferior colliculus forms the primary site of convergence for these streams, and is a critical stage in the transformation from simple auditory responses to complex auditory objects (Pickles, 2008, p. 183). The inferior colliculus then projects up to the medial geniculate body, which acts as an auditory relay between the inferior colliculus and auditory cortex.

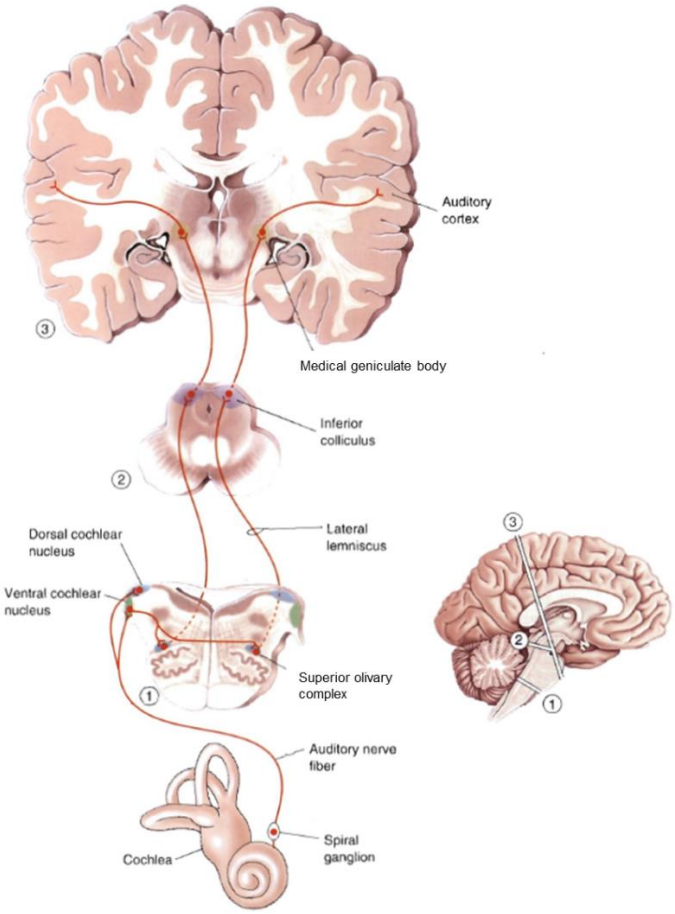

Figure 2.2: Central Auditory System

A diagram of the central auditory system with simplified connections. Adapted from Bear et al. (2007).

2.1.3 Auditory Cortex

The auditory cortex is located in the superior portion of the temporal lobe (Figure 2.3). It can be subdivided into three areas, i.e., the core, belt, and parabelt, the functions of which tend to show a progressive increase in complexity from inside out. The core is located deep within the lateral sulcus and receives input from the medial geniculate body. Like the cochlea, the core is tonotopically organised (Pickles, 2008, p. 207). The belt is a narrow band of cortex that surrounds the core, that also receives input from the medial geniculate body as well as the core (Pickles, 2008, p. 209). Belt regions are highly interconnected and can also be tonotopically organised. The parabelt receives input from the belt, and is connected to several areas of the frontal, parietal and temporal lobes. These areas tend to be involved in high-level, e.g., speech, processing and are beyond the scope of this thesis.

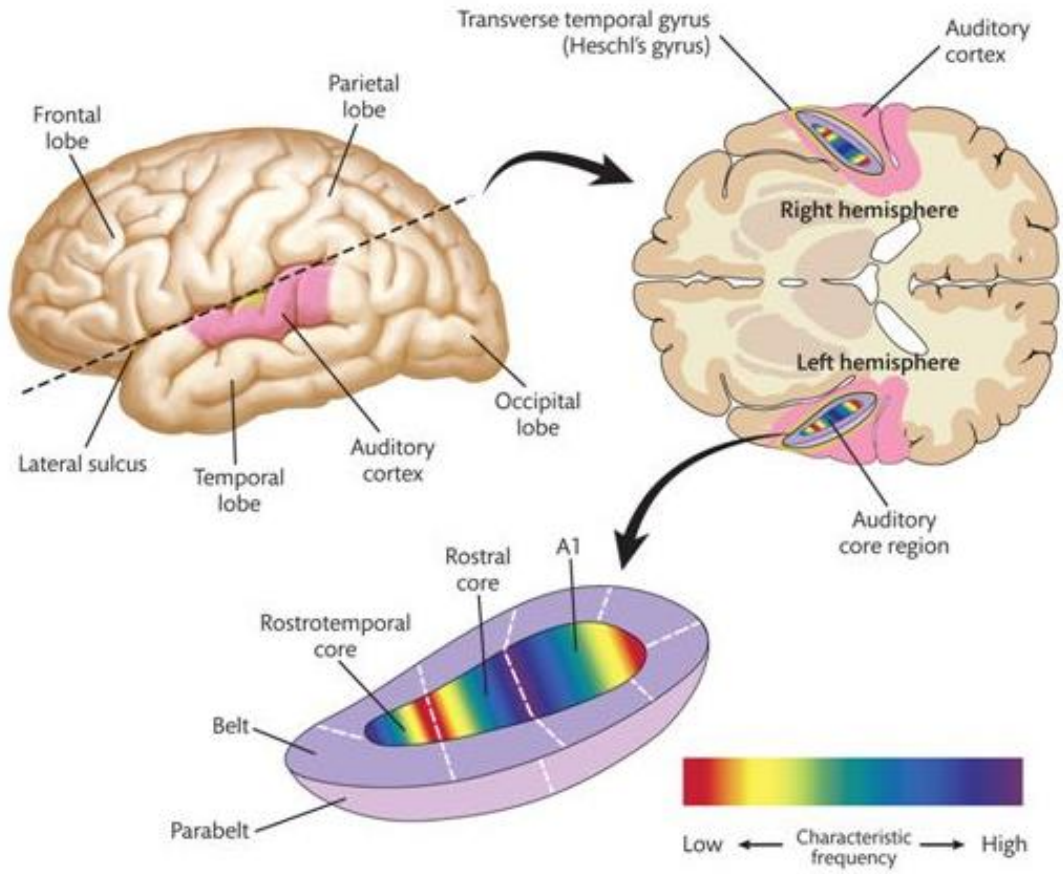

Figure 2.3: Auditory Cortex

A diagram of the auditory cortex including the relationships between the core, belt, and parabelt areas.

2.2 Pathophysiology of Human Auditory Processing

2.2.1 Conductive Hearing Loss

Conductive hearing loss refers to hearing loss due to some abnormality in the outer- or middle-ear. In these parts of the hearing system, sound is transferred through movement, e.g., the movement of air along the ear canal or the movement of the tympanic membrane and ossicles, and so the cause of conductive hearing loss tends to be something that impedes this movement in some way, e.g., the ear canal may be blocked, the tympanic membrane damaged, or the ossicles immobilised due to ossification, etc. This type of hearing loss can be well compensated by hearing aids, and in more severe cases surgical interventions, e.g., by replacing the stapes with a prosthesis (Pickles, 2008, p. 309).

2.2.2 Sensorineural Hearing Loss

Sensorineural hearing loss refers to hearing loss due to some problem arising in the cochlea or auditory nerve. While this can be caused by a benign tumour around the sheath of the auditory nerve known as an acoustic neuroma, it is most commonly an issue with the hair cells of the cochlea. These issues can be caused by acoustic trauma, drugs, infections, or may be congenital or due to old age (Pickles, 2008, p. 309). Unfortunately, hearing aids tend to be of limited use for this type of hearing loss, although in more severe cases cochlear implants—electronic prostheses that aim to replicate the mechanical-to-electrical transduction of the cochlea—can provide a limited sensation of hearing to those who otherwise would have none.

2.2.3 Hidden Hearing Loss

Hidden hearing loss refers to hearing loss typically associated with a difficulty understanding speech-in-noise, that presents despite normal PT audiograms, i.e. it is “hidden” from such measures. Kujawa and Liberman (2009) suggest that this may be due to cochlear synaptopathy, i.e., degeneration of the synapses between hair cells and auditory nerve fibres, predominantly in high-threshold fibres. This would mean that while threshold detection in quiet—which is what is tested with PTA—would remain unaffected, threshold detection in noise—for which high-threshold fibres are essential—could become much worse. Musiek et al. (2018) on the other hand contend that central auditory processing disorder—a deficit in high-level auditory processing in the central nervous system—is the most likely cause of hidden hearing loss.

While this type of hearing loss can remain “hidden” with PTA, it can be revealed using appropriate high-level measures, such as those involving dichotic listening, speech-in-noise, etc.

2.3 Electrophysiology of Low-Level Human Auditory Processing

2.3.1 Electroencephalography

Once the mechanical-to-electrical transduction has taken place in the cochlea, auditory processing becomes electrical—or more specifically, electrochemical—in nature. Signals propagate along the auditory pathway in the form of action potentials—all-or-nothing electrochemical events that depolarise sections of nerve fibres, and neurons, provided that sufficient membrane potentials have been met. These electrical signals can be recorded using electrodes placed on the surface of the scalp using a technique referred to as EEG and can provide objective measures of auditory processing.

The connections between neurons are known as synapses, and it is typically the electrical fields generated by synchronous postsynaptic activity from many neurons that is what is recorded with EEG. As EEG electrodes are placed on the surface of the scalp, these electrical fields need to pass through several anatomical layers, e.g., cerebrospinal fluid, bone, and skin, before reaching the electrode surface. This leads to an attenuation of the signal, particularly in-contrast to other larger electrophysiological signals such as those of the electrooculogram (EOG), e.g., as a result of blinking or eye movement, or electromyogram (EMG), e.g., as a result of muscle movement, which are often also recorded along with the EEG. It can also lead to spatial smearing at the scalp—particularly as a result of passing through the skull (Srinivasan et al., 1996)—meaning that the activity recorded at a single electrode can comprise a mixture of underlying sources (Makeig et al., 1996).

However, what EEG lacks in SNR and spatial resolution, it makes up for in temporal resolution—which is in the order of milliseconds. This, combined with its relatively low-cost and portability, has made it a very suitable method for investigating human auditory processing.

2.3.2 Electrophysiological Measures of Low-Level Human Auditory Processing

There are numerous measures of low-level human auditory processing that can be recovered with EEG. These can be categorised by latency and time-course (Table 2.1). The measures that are of most relevance to this work are the early, middle, and late transient responses (Picton, 2010, p. 5), specifically the auditory brainstem response (ABR), the middle-latency response (MLR), and the late auditory evoked potential (LAEP), which together comprise the multiple-latency AEP. The reason that these responses are so relevant to the current work is that they can be considered as special cases of the TRF (see Chapter 3 for details). Briefly, the AEP is the auditory system response to (quasi-)discrete stimuli while the TRF represents the auditory system response to any stimuli—including (quasi-)discrete stimuli.

Table 2.1: Low-Level Auditory Responses

Latency	**Time-Course**		
	Transient	**Steady-State**	**Sustained**
Early (0–10 ms)	Cochlear Nerve Compound Action Potential (CAP); Auditory Brainstem Response (ABR)	Cochlear Microphonic (CM); Frequency Following Response (FFR); Fast Auditory Steady-State Response (> 70 Hz ASSR)	Summating Potential (SP); Pedestal of Frequency Following Response
Middle (10–50 ms)	Middle-Latency Response (MLR)	Auditory Steady-State Response (ASSR)	-
Late (50–1000 ms)	Late Auditory Evoked Potential (LAEP); Mismatch Negativity (MMN); Processing Negativity; Late Positive Waves	Slow Auditory Steady-State Response (< 30 Hz ASSR)	Cortical Sustained Potential (SP); Contingent Negative Variation (CNV)

2.3.3 Auditory Evoked Potential

The ABR refers to the transient response recorded from the auditory brainstem, i.e., much of the early central auditory system, between 0–10 ms after the onset of a brief sound (Figure 2.4). It comprises a series of 6 or 7 waves, usually numbered with roman numerals, i.e., I-VII, according to the convention of Jewett and Williston (1971). As some of the waves are quite variable in terms of amplitude and identifiability, it is usually waves I, III, and V that are

evaluated (Burkard et al., 2007, p. 230). While it can be helpful to think of each wave as representing a distinct stage of neural processing along the auditory pathway, most waves—apart from wave I which is generated solely by the auditory nerve—are generated by multiple neural sources.

The MLR refers to the transient response recorded from the upper brainstem and thalamic nuclei, i.e., the late central auditory system and auditory cortex, between 10–50 ms after the onset of a brief sound (Figure 2.4). It comprises a series of 5 waves—2 positive and 3 negative—i.e., No-Po-Na-Pa-Nb, following the convention of Goldstein and Rodman (1967). N and P denote the polarity of the waves when recorded at the vertex. The shape and amplitude of these waves are sensitive to stimulus, subject, and recording factors.

The LAEP refers to the transient response recorded from the auditory cortex, between 50–1000 ms after the onset of a brief sound. It comprises a series of 4 waves—2 positive and 2 negative—i.e., P1-N1-P2-N2, following the convention of Williams et al., (1962). The shape and amplitude of these waves are also sensitive to stimulus, subject, and recording factors.

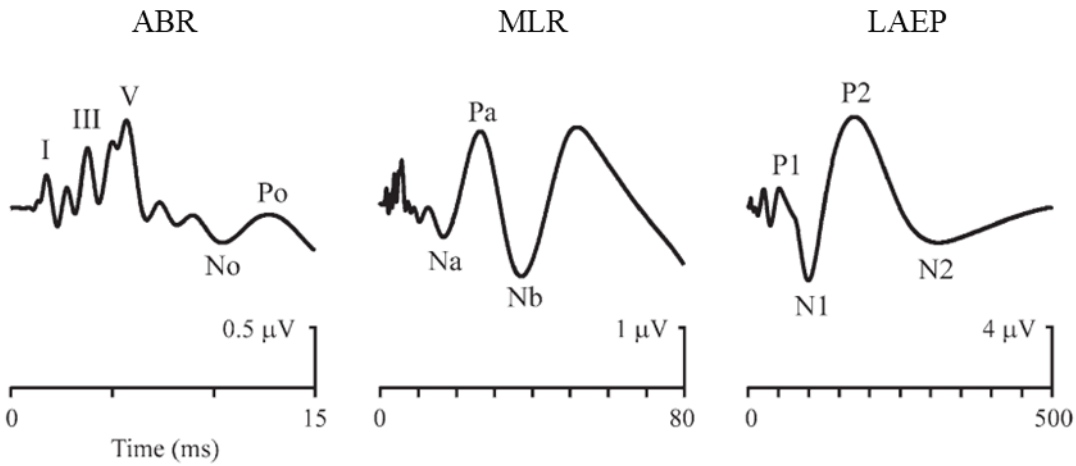

Figure 2.4: ABR, MLR, and LAEP

Example ABR, MLR, and LAEP with labelled waves. Please note the different time and amplitude scales used. Adapted from Picton (2013).

2.3.4 Recovering Auditory Evoked Potentials

Time-domain averaging (TDA) has long been the canonical approach for recovering AEPs from EEG (Geisler et al., 1958). In TDA, the response to each individual (quasi-)discrete stimulus is combined and averaged, with the assumption that if the noise in the response is random, it should average out to zero while the signal, i.e., the evoked potential—which is not random—should remain intact, thus increasing the SNR of the signal (Figure 2.5). It is worth

noting, however, that this increase in SNR is proportional to the square root of the number of averages (Hall, 1992, p. 82), and so comes with diminishing returns. The exquisite temporal resolution afforded by TDA has produced the canonical ABRs, MLRs, and LAEPs that have come to be so heavily studied. As such, TDA has been instrumental in advancing our knowledge of the human auditory system.

One shortcoming of the TDA approach, however, is that one is typically restricted to using (quasi-)discrete stimuli. Another shortcoming is that data collection can be quite slow as the maximum rate at which stimuli can be presented is limited. This is because a succeeding stimulus typically cannot be presented until the response to the preceding stimulus has ended, as otherwise a confusing overlap of responses would occur. Traditionally, this has meant that separate sets of specialised stimuli and resulting neural data are required to assess each auditory latency of interest, i.e., it has not been possible to recover multiple-latency responses using just one set of stimuli and resulting EEG.

Several methods have permitted faster presentation rates to be used by attempting to account for the resulting overlapping responses. These include maximum length sequence (MLS or m-sequence) deconvolution (Eysholdt and Schreiner, 1982), the adjacent response (ADJAR) technique (Woldorff, 1993), continuous loop averaging deconvolution (CLAD; Delgado and Ozdamar, 2004), q-sequence deconvolution (QSD; Jewett et al., 2004), randomized stimulation and averaging (RSA; Valderrama et al., 2012), multiple-rate steady-state deconvolution (MSAD; Wang et al., 2013), and least-squares (LS) deconvolution (Bardy et al., 2014). While each new method has contributed certain improvements, e.g., increased flexibility—in terms of the stimuli that can be used—or efficiency, simplified application or increased robustness, the presentation rate limit has remained problematic and one is still typically constrained to using restricted sets of stimuli. That is not to say that these approaches are without merit, however, having shown particular promise in the study of adaptive mechanisms (Burkard et al., 1990; Lasky, 1997) and in the diagnosis of certain pathologies (Tanaka et al., 1996; Jiang et al., 2000). While other methods focused on the use of slower presentation rates optimised for the recording of dual- or multiple-latency responses have been more successful (Bidelman, 2015; Kohl and Strauss, 2016), they have also necessarily resulted in certain compromises.

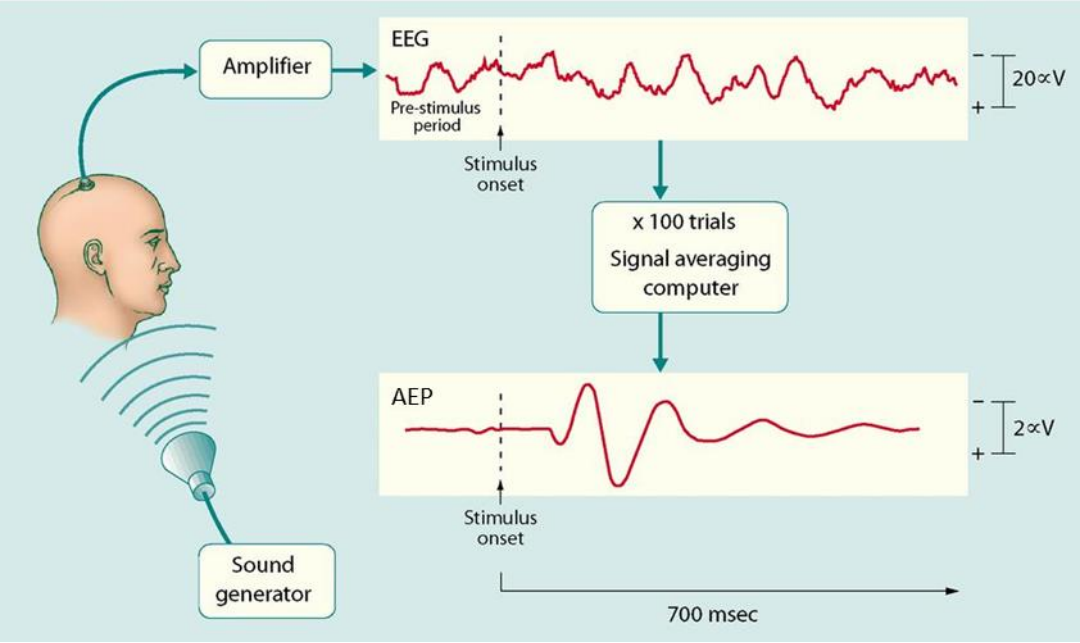

Figure 2.5: Time-Domain Averaging

In Time-Domain Averaging (TDA), the auditory response to a presented sound is recorded using EEG electrodes, amplified, and stored. This process is typically repeated hundreds to thousands of times—depending on the SNR of the response—and the responses combined and averaged. If the noise in the response is random, it should average out to zero while the signal, i.e. the evoked potential—which is not random—should remain intact, thus increasing the SNR of the signal.

Chapter 3. Temporal Response Function Estimation

In this chapter, the TRF estimation approach is introduced. This is followed by a description of its theoretical/mathematical basis, including definitions of relevant terminology, equations, and procedures. Some nomenclature associated with different types of stimulus representations are then defined, followed by a discussion on how to interpret the resulting TRFs. Finally, the relationship between the TRF and AEP is outlined.

3.1.1 Introduction

In recent years, several methods that facilitate the use of a wider variety of stimuli have been proposed. These methods typically involve fitting mathematical models that describe the relationship between a sensory input and a neural response—an approach referred to as system identification (see Wu et al., 2006 for a review). A central feature of such models has been a linear receptive field stage that seeks to account for some of the neural response as a linear weighted sum, i.e., a linear filter, of particular features of the sensory input, e.g., the amplitude envelope. In neural spiking models, this linear filtering stage is typically just one of several stages, e.g., linear, nonlinear, and Poisson, that seek to capture how variations in the stimulus are reflected in spike trains (Chichilnisky, 2001). However, with more macroscopic data like functional magnetic resonance imaging (fMRI; Boynton et al., 1996) or EEG (Crosse et al., 2016), this linear filtering stage often represents the entirety of the model. While the human brain is not linear, this assumption can be reasonable in certain cases, e.g., when dealing with macroscopic data (Boynton et al., 1996). While the use of non-linear modelling approaches can lead to slight increases in modelling performance (Power et al., 2011), the resulting increase in

complexity and decrease in interpretability has led to linear modelling approaches being favoured.

One explicit use of this linear modelling approach has been the fitting of TRFs to describe how EEG is affected by variations in visual (Gonçalves et al., 2014) or—as in this case—auditory stimuli (Lalor et al., 2009). With TRF estimation, the assumption is that the instantaneous output EEG at channel n of N, $y(t,n)$, sampled at times $t = 1 \ldots T$, comprises the linear convolution of an input stimulus feature, $x(t-\tau)$, with an unknown system response, $w(\tau,n)$, i.e., the TRF, plus noise (Figure 3.1):

$$y(t,n) = \sum_{\tau} w(\tau,n)x(t-\tau) + \varepsilon(t,n)$$

where τ represents the **range of time-lags** over which the TRF is estimated, and $\varepsilon(t)$ represents the residual EEG not explained by the model (Crosse et al., 2016). The range of time-lags used to derive a TRF is typically similar to that used to recover an AEP, although the interpretation of their timing is somewhat different. For example, the value of an AEP at 100 ms describes the average response in the EEG 100 ms after stimulus onset, whereas the value of a TRF at 100 ms describes how a change in the stimulus feature will affect the EEG 100 ms later (Lalor et al., 2009).

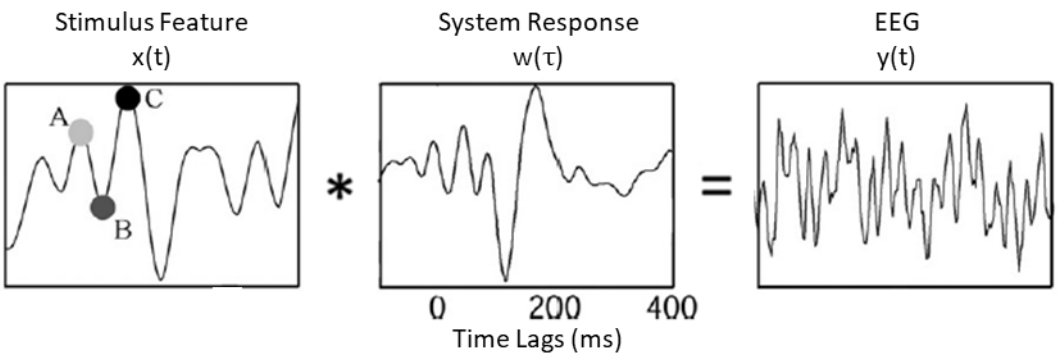

Figure 3.1: TRF Estimation

With TRF estimation, the assumption is that the output EEG, $\boldsymbol{y(t)}$, consists of the convolution of a particular input stimulus feature, $\boldsymbol{x(t)}$, with an unknown system response, i.e., the TRF, $\boldsymbol{w(\tau)}$, plus noise (Lalor et al., 2009). Given the known stimulus feature and the measured EEG, the TRF can be derived (for example) by performing regularised linear least-squares estimation with ridge regression (Crosse et al., 2016). Adapted from Lalor et al. (2009).

3.1.2 Forward Models

TRFs that describe the mapping from the stimulus feature to the EEG are known as forward models. Forward models are derived by minimising the mean-squared error (MSE) between the predicted EEG, $\hat{y}(t,n)$, and measured EEG, $y(t,n)$:

$$\min \varepsilon(t,n) = \sum_{t} [\hat{y}(t,n) - y(t,n)]^2$$

In practice this is done using linear regression:

$$w = (X^T X)^{-1} X^T y$$

where y is a matrix of EEG data with channels arranged column-wise, X is the **design matrix** or lagged time series of the stimulus feature $x(t)$:

$$X = \begin{bmatrix} x(1-\tau_{min}) & x(-\tau_{min}) & \cdots & x(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & x(1) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & x(1) \\ x(T) & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & x(T) & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x(T) & x(T-1) & \cdots & x(T-\tau_{max}) \end{bmatrix}$$

where τ_{min} and τ_{max} are the minimum and maximum time lags (in samples) respectively, $X^T X$ is the **autocovariance matrix**—which can be problematic to invert—and $X^T y$ is the **cross-covariance matrix** (Crosse et al., 2016). In X, each time lag is arranged column-wise and non-zero lags are padded with zeros to ensure causality (Mesgarani et al., 2009). The **window** or **window of support** over which the TRF is calculated is defined as $\tau_{window} = \tau_{max} - \tau_{min}$, and so the dimensions of X are $T \times \tau_{window}$, although a column of ones is also concatenated to the left of X to include the y-intercept in the regression model (Crosse et al., 2016). y has dimensions $T \times N$ and the resulting TRF, w, has dimensions $\tau_{window} \times N$, where each column represents the mapping between the stimulus feature and EEG at a different channel (Crosse et al., 2016).

3.1.3 Backward Models

TRFs that describe the mapping from the EEG back to the stimulus feature are known as backward models and offer a complementary way to investigate how stimulus features are encoded in EEG (Crosse et al., 2016). With forward modelling, in-effect, separate stimulus-

response mappings are derived for each EEG channel, whereas with backward modelling, all of the available data is exploited simultaneously when deriving reverse stimulus-response mappings (Crosse et al., 2016). This makes backward models more sensitive to small differences between EEG channels that are highly correlated with each other—as is often the case with EEG (Crosse et al., 2016)—although this does come at the cost of direct neurophysiological interpretability (Haufe et al., 2014).

Like forward models, backward models are derived by minimising the MSE between the reconstructed stimulus feature, $\hat{x}(t)$, and actual stimulus feature, $x(t)$:

$$\min \varepsilon(t) = \sum_{t} [\hat{x}(t) - x(t)]^2$$

Again, in practice, this is done using linear regression:

$$g = (Y^T Y)^{-1} Y^T x$$

where x is a column-wise vector or matrix containing the stimulus feature—depending on whether it is a univariate or multivariate feature, e.g., an envelope or spectrogram—and Y is the lagged time series of the EEG matrix, y. For simplicity, for a single channel system:

$$Y = \begin{bmatrix} y(1-\tau_{min},1) & y(-\tau_{min},1) & \cdots & y(1,1) & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & y(1,1) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & y(1,1) \\ y(T,1) & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & y(T,1) & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & y(T,1) & y(T-1,1) & \cdots & y(T-\tau_{max},1) \end{bmatrix}$$

The range of time-lags, τ, used here would typically be similar to that used in forward modelling except in the reverse direction, as it is effectively mapping backwards in time (Crosse et al., 2016). So instead of the lags ranging from -100 to 400 ms for example, they would range from -400 to 100 ms. The dimensions of Y for a single channel system are $T \times \tau_{window}$, although a column of ones is also concatenated to the left of Y to include the y-intercept in the regression model. This can be extended to an N-channel system by replacing each column with N columns, each representing a separate channel (Crosse et al., 2016). In this case the dimensions of Y would be $T \times N\tau_{window}$, although N columns of ones would then also then be concatenated to the left of Y. x can have dimensions $T \times 1$ or $T \times F$ depending on whether it contains a univariate or multivariate stimulus feature, with the resulting model, g, then having dimensions $N\tau_{window} \times 1$ or $N\tau_{window} \times F$, respectively (Crosse et al., 2016).

3.1.4 Regularisation

Two issues that can arise when deriving TRFs involve the inversion of ill-conditioned matrices, and **overfitting**. Matrix inversion is prone to numerical instability when solved with finite precision, and so small changes in the autocovariance matrix, $X^T X$, can cause large changes in the TRF if the former is ill-conditioned (Crosse et al., 2016). Overfitting occurs when a TRF has become optimally fit for a particular dataset but does not generalise well to unseen data. This is often because the TRF has also been fit to the "noise" in the data it has been trained on, which is unlikely to be present in any other (Crosse et al., 2016). Both issues can be improved using a method known as regularisation, i.e., the introduction of a bias or "smoothing" term to reduce the variance in the TRF.

In practice, regularisation can be carried out by weighting the diagonal of $X^T X$ before inversion—a method known as ridge regression:

$$w = (X^T X + \lambda I)^{-1} X^T y$$

where λ is the bias term (**regularisation parameter**) and I is the identity matrix. This form of regularisation enforces a smoothness constraint on the TRF by penalising TRF values as a function of their distance from zero (Crosse et al., 2016). Another option is to quadratically penalise the difference between each two neighbouring terms of the TRF, a method known as Tikhonov regularisation (Tikhonov, 1963):

$$w = (X^T X + \lambda M)^{-1} X^T$$

where:

$$M = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}$$

Unlike ridge regression, Tikhonov regularisation preserves the signal amplitude when smoothing the TRF, and so often produces a better estimate (Crosse et al., 2016; Wong et al., 2018). The optimal value for λ is typically determined using cross-validation, as will be discussed in the next section.

3.1.5 Model Fitting Procedure

The first step when deriving TRFs is to fit a separate model for each of M trials. One trial is then typically chosen to be left-out, i.e., to be used as the validation/test set, with the remaining M-1 trials to be used as the training set. An average model is then attained by averaging over the single-trial models in the training set, before being convolved with the stimulus feature associated with the validation/test set to predict its EEG response. Model performance is then assessed by quantifying how accurately the predicted EEG response matches the recorded EEG response in the validation/test set. This measure is often attained using Pearson's correlation coefficient and is referred to as **prediction accuracy**. This entire process—which is referred to as cross-validation—is then repeated M-1 times such that each trial is left-out of the training set once. The overall model performance can then finally be determined by averaging over the individual model performances for each trial (Crosse et al., 2016).

This procedure can also be carried out in the backward direction, where instead of trying to predict the EEG response of the left-out trial, one is trying to reconstruct the stimulus feature used to elicit its EEG response (Crosse et al., 2016). In this case, model performance can be assessed by quantifying how accurately the reconstructed stimulus feature matches the presented stimulus feature associated with the validation/test set. This measure is again often attained using Pearson's correlation coefficient and is referred to as **reconstruction accuracy**. Because the models are fit using regularisation, it is also necessary to determine the optimal lambda value. This is done by repeating the cross-validation procedure for each of several lambda values and then choosing the value that results in the best prediction/reconstruction accuracy.

3.1.6 Stimulus Representation

As mentioned, the main advantage of TRF estimation is that it can be used with almost any stimulus. Indeed, using this approach, Lalor et al. (2009) showed that it is possible to derive LAEPs in response to quasi-periodic tone-bursts, AMTs, and AM BBN, and Lalor and Foxe (2010) showed that it is possible to derive LAEPs in response to continuous natural speech. Recently, Maddox and Lee (2018) modified this approach and using it showed that it is possible to derive responses to continuous natural speech from the entire auditory pathway simultaneously. This finding provides a long-sought solution to the issue of deriving responses

from multiple latencies simultaneously, i.e., using just one set of stimuli and resulting EEG, and does so while permitting a wider variety of stimuli to be used.

Given that the modelling approach and EEG are typically fixed, the choice of stimulus feature and the way in which it is represented—collectively referred to as the **stimulus representation**—plays an important role in determining the resulting model. For example, depending on whether one uses a global measure of amplitude change like the envelope, or a multivariate binary representation of phoneme activity, such as the phoneme representation introduced by Di Liberto et al. (2015a), one can interrogate very different aspects of auditory processing. How these stimulus representations are generated, and what they can be used for will be discussed in more detail throughout the thesis. First, however, it would be useful to define a nomenclature with which to describe them.

Feature

'Feature' refers to some property of the original stimulus that is encapsulated by the stimulus representation. Classically the envelope has been used (Lalor and Foxe, 2010) but other features such as the raw audio signal (Maddox and Lee, 2018), phonemes, phonetic features (Di Liberto et al., 2015), and semantic dissimilarity (Broderick et al., 2018) have also proven useful. The choice of feature is dependent on the attribute and/or result of interest.

Form

'Form' refers to the way in which the stimulus representation is expressed. For example, stimulus representations are typically expressed in their voltage form, i.e., the form in which they are stored in the audio file. However, as electrophysiological responses tend to vary in proportion to the log of the stimulus amplitude, for example (Aiken and Picton, 2008), they could also be represented in their in their log or "sound pressure level" (SPL) form (as in Chapter 6).

Compensation

'Compensation' refers to some modification of the stimulus representation to account for known electrophysiological properties of the auditory system. For example, this could include the application of a gammachirp filter bank to compensate for the cochlear-neural time delay (as in Chapter 4), or the encoding of the stimulus representation as MP3 or AAC, to compensate for the masking effects of the auditory system. Such compensations are more or less appropriate

depending on the goal, e.g., the application of a gammachirp filter bank would obviously be inappropriate if trying to study the cochlear-neural time-delay.

Calculation

‘Calculation’ refers to the calculation of some higher-order feature from the original stimulus feature. For example, this could be the positive or negative half-wave rectified audio signal (Maddox and Lee, 2018), the full-wave rectified audio signal, the positive half-wave rectified first derivative (onset) of the envelope (Hertrich et al., 2012; Fiedler et al., 2016; as in Chapter 6), the negative half-wave rectified first derivative (offset) of the envelope, or the full-wave rectified first derivative of the envelope.

Binning

‘Binning’ refers to the binning of the stimulus representation based on chosen attributes. The stimulus representation could be frequency-binned (FB), i.e., a spectrogram, amplitude-binned (AB; as in Chapter 6), or frequency- and amplitude-binned (FAB).

Value

‘Value’ refers to the value given to the binned features. For example, these could be their original values, or perhaps they may be categorical such as phonetic features and so be represented as binary, or in the case of semantic dissimilarity, the value of the feature corresponds to the semantic dissimilarity at that point.

The specific combination of these parameters can be used to describe the stimulus representation. For example, a raw audio signal stored in a WAV file or similar, could be referred to simply as a voltage signal, while an audio signal that has been passed through a gammachirp filter bank, had its envelope extracted, transformed to SPL, and amplitude-binned, could be referred to as an AB gammachirp SPL envelope. There are certain situations where parameters are implicit, but for others perhaps this format could be useful.

3.1.7 Interpretation

As alluded to earlier, prediction and reconstruction accuracies can be used as dependent measures. For example, if one has a hypothesis about how a stimulus feature might be represented in the brain, one can test that hypothesis by training a model using a stimulus

representation that is exemplary of that hypothesis, and then evaluating its fit using the resulting prediction or reconstruction accuracies. If the chosen feature is indeed represented in the brain in the manner hypothesised, it should result in high prediction or reconstruction accuracies. What is defined as "high" in this instance is relative, as the actual values tend to be quite low. This is because the ratio of EEG associated with the stimulus representation to EEG not associated with the stimulus representation also tends to be quite low, and so, prediction and reconstruction accuracies need to be interpreted in this light. Analysis of the model parameters themselves can also provide some insight. For example, one can examine and compare the TRF values at different channels and time-lags to determine which cortical regions may be contributing to the response, and when. These approaches are used extensively throughout the thesis.

3.1.8 Relationship to AEP

As mentioned, deriving a TRF is somewhat analogous to recovering an AEP. An AEP is an auditory system response to a specialised set of discrete stimuli, while a TRF represents a linear best-fit to the auditory system response to any stimuli—including discrete. Indeed, a TRF derived to discrete stimuli can be equivalent to an AEP, provided the window of support, i.e., the range of time-lags over which the TRF is estimated, is shorter than the stimulus onset asynchrony (SOA; Lalor et al., 2009), i.e., the time between stimulus onsets, and the stimuli are represented as impulses—given that the model only reflects the response to what is characterised by the stimulus representation. This cannot be true of a TRF derived to continuous stimuli, however, as the auditory system response will change depending on the stimulus. Therefore, the AEP can be considered a special case of the TRF.

Chapter 4. Indexing the Human Auditory Processing Hierarchy: Considerations of Stimulus Type, Stimulus Representation, and Computational Efficiency

In this chapter, two approaches for indexing low-level processing along the auditory pathway are introduced over two experiments: Experiment 1 and 2. Experiment 1 is an initial exploratory attempt to derive multiple-latency responses to high-rate click trains. Experiment 2 is a more thorough attempt to derive multiple-latency responses to AM BBN, while introducing a novel efficient TRF estimation approach. A manuscript on Experiment 2 is in preparation at the time of writing for submission to the Journal of the Acoustical Society of America.

Experiment 1

4.1 Introduction

As mentioned in Chapter 2, TDA has long been the canonical approach for recovering AEPs from EEG (Geisler et al., 1958) and has been instrumental in advancing our knowledge of the human auditory system. While the restriction to the use of (quasi-)discrete stimuli can be seen as a disadvantage, it has also resulted in one set of prototypical responses becoming very heavily studied, i.e., the ABR, MLR, and LAEP. However, data collection with TDA can be relatively slow as the maximum rate at which stimuli can be presented is limited—as otherwise a confusing overlap of responses would occur. Traditionally, this has meant that separate sets of specialised stimuli and resulting EEG are required to assess each auditory latency of interest,

i.e., it has not been possible to recover multiple-latency responses using just one set of stimuli and resulting EEG. Facilitating the efficient recovery of these canonical responses could have wide ranging benefits for both research and clinical applications where these responses remain instrumental.

As also mentioned in Chapter 2, several methods have permitted faster presentation rates to be used by attempting to account for the overlapping responses. These include MLS deconvolution (Eysholdt and Schreiner, 1982), the ADJAR technique (Woldorff, 1993), CLAD (Delgado and Ozdamar, 2004), QSD (Jewett et al., 2004), RSA (Valderrama et al., 2012), MSAD (Wang et al., 2013), and LS deconvolution (Bardy et al., 2014). While these methods have enabled recovery from overlapping responses, the presentation rate limit has remained problematic. Other methods focused on the use of slower presentation rates, optimised for the recording of dual- or multiple-latency responses (Bidelman, 2015; Kohl and Strauss, 2016) have been more successful, but have also necessarily resulted in certain compromises.

As TRF estimation has been shown to be effective in the derivation of responses to continuous stimuli (Lalor et al., 2009; Lalor and Foxe, 2010), we were curious as to whether it could be used to improve upon previous efforts to permit faster discrete presentation rates. To the author's knowledge, this had not been attempted before, i.e., it was not known at the time of this experiment's inception that a similar approach had been proposed by Bardy et al., (2014). Hence, in this experiment we aim to investigate the utility of TRF estimation in the derivation of canonical multiple-latency responses to high-rate click trains. The effectiveness of this approach will be determined through comparisons with canonical AEPs to suitable-rate click trains, derived using TRF estimation.

4.2 Materials and Methods

4.2.1 Subjects

8 subjects aged 23–27 years participated in this study; 6 were male. All subjects had self-reported normal hearing. The protocol for this study was approved by the Ethics Committee of the Health Sciences Faculty at Trinity College Dublin, Ireland, and all subjects gave written informed consent.

4.2.2 Stimuli

Alternating 100 µs click trains with uniformly distributed ±25% temporal jitter, were presented at rates of 58.4, 21.9, 7.3, and 1.0 Hz (Figure 4.1). These specific rates were chosen because they are suitable—in terms of having sufficiently long SOAs—for evoking canonical ABRs—with some degradation of the earlier waves—ABRs—with little degradation of the earlier waves—MLRs, and LAEPs, respectively.

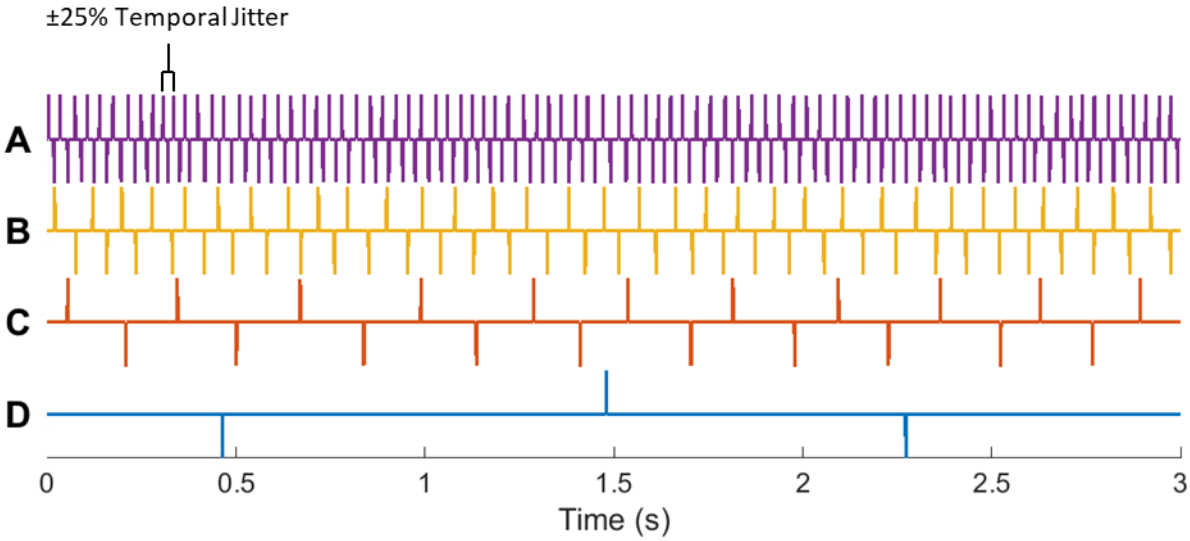

Figure 4.1: Example Segments of the Stimuli Used in This Experiment

A–D *– Alternating 100 µs click trains with uniformly distributed ±25% temporal jitter, presented at 58.4, 21.9, 7.3, and 1.0 Hz, respectively.*

4.2.3 Experimental Procedure

Subjects were presented with 10 repetitions each of the same 60 s long 58.4 Hz, 21.9 Hz, and 7.3 Hz click trains, and 7 repetitions of the same 60 s long 1.0 Hz click train, as they sat in a comfortable chair, in a quiet, darkened room, and watched a muted, subtitled film presented on a laptop computer. Fewer repetitions were required for the 1.0 Hz click train as the responses it is intended to elicit, i.e., LAEPs, have the highest SNR. Subjects were asked not to attend to the auditory stimuli, which were presented monaurally to their right ear. The clicks were presented at 98 dB peak-to-peak equivalent SPL (peSPL), i.e., the peak-to-peak amplitude of the clicks matched that of a 1 kHz PT presented at 98 dB SPL. This equates to ~70 dB nHL, i.e., 70 dB above the typical threshold for click stimuli for normal hearing subjects, based on the 28 dB peak-to-peak reference equivalent threshold SPL (peRETSPL), i.e., the reference threshold for click stimuli for normal hearing subjects, given for the Sennheiser HDA 200

headphones (ISO 389-6:2007). The click trains were presented using Sennheiser HD 650 headphones, via Presentation software from Neurobehavioral Systems (http://www.neurobs.com). The stimulus presentation order, i.e., for each 60 s long stimulus, was pseudorandomised to minimise any potential order effects.

4.2.4 EEG Acquisition

34 channels of EEG data were recorded at 16384 Hz (analog -3 dB point of 3276.8 Hz), using a BioSemi ActiveTwo system (http://www.biosemi.com). 32 cephalic electrodes were positioned according to the standard 10-20 system, with another 2 electrodes located over the left and right mastoids. Triggers indicating the start of each 60 s trial were presented using Neurobehavioral Systems Presentation software for synchronous recording along with the EEG.

4.2.5 EEG Preprocessing

The EEG data were first resampled to 128 Hz using the *decimate* function in MATLAB (http://www.mathworks.com). The *decimate* function incorporates an 8th order low-pass Chebyshev Type I infinite impulse response (IIR) anti-aliasing filter. This filter was implemented using the *filtfilt* function, ensuring zero phase distortion and in effect doubling the order of the filter. A 1st order high-pass Butterworth filter was then applied with a cutoff frequency of 1 Hz, also using the *filtfilt* function. Bad channels were determined as those whose variance was either less than half or greater than twice that of the surrounding 2–4 channels, depending on location. These were then replaced through spherical spline interpolation using EEGLAB (Delorme and Makeig, 2004). Finally, the data were rereferenced to the average of the mastoids, separated into trials based on the triggers provided, and z-scored.

4.2.6 Temporal Response Function Estimation

Responses were derived using TRF estimation, and were implemented via the mTRF Toolbox (Crosse et al., 2016) using a forward modelling approach (see Chapter 3 for details). Baseline correction was performed on each subject's average TRF by subtracting the mean value between -20 and 0 ms, before being combined to form the grand average response.

4.3 Results

First, grand average click train LAEPs were derived in response to each of the four click train stimuli, and the resulting waveforms plotted—stacked on top of one another—on the left-hand side of Figure 4.2. It is clear from these plots that the morphologies (shapes) of the 58.4, 21.9, and 7.3 Hz click train LAEPs are quite different from that of the canonical 1.0 Hz click train LAEP. This—albeit qualitative—comparison indicates that it is not possible to derive canonical multiple-latency responses to high-rate click trains using TRF estimation, and that any efforts to derive MLRs or ABRs would be moot.

These results concur with previous efforts to derive multiple-latency responses using high-rate discrete stimuli. Indeed these results effectively mirror those of high-rate chirp train LAEPs derived using CLAD (Holt and Özdamar, 2016; right-hand side of Figure 4.2). The high degree of similarity between these two sets of responses, suggests that this is due to some fundamental feature of the auditory system rather than being a technical limitation inherent to any one approach.

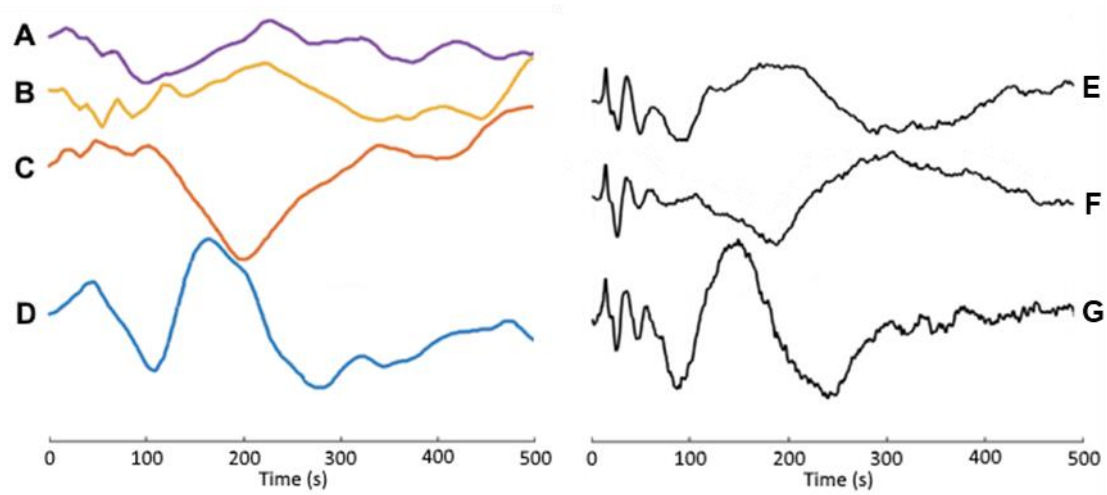

Figure 4.2: Grand Average Click Train vs. Grand Average Chirp Train LAEPs

A–D *– Grand average LAEPs derived to the 58.4, 21.9, 7.3, and 1.0 Hz click trains used in this experiment, respectively. E–G – Grand average LAEPs derived by Holt and Özdamar (2016) to 20.0, 7.0, and 1.0 Hz chirp trains, respectively.*

4.4 Discussion

Despite numerous attempts, from several groups, over many years, the presentation rate limit has remained problematic. As is apparent from Figure 4.2, the morphologies of the 58.4, 21.9, and 7.3 Hz click train LAEPs are quite different from that of the canonical 1.0 Hz click train

LAEP. This is likely due to response adaptation at the higher rates, as has been seen with other similar approaches, e.g., CLAD (Holt and Özdamar, 2016). Adaptation in this context refers to changes in response amplitude as a function of SOA. This manifests as changes in response morphology because LAEP waves tend to comprise multiple components and these components tend to be differently affected by SOA (Lü et al., 1992; Sams et al., 1993). Therefore, it is highly unlikely that any approach employing high-rate discrete stimuli would ever be able to derive canonical LAEPs—and thus canonical multiple-latency responses—as the morphology of these responses are fundamentally dependent on the SOAs of the stimuli used to elicit them.

While it may not be possible to derive canonical LAEPs using high-rate discrete stimuli, it is possible to recover LAEPs that display the effects of adaptation. Such responses could be useful in the study of adaptive mechanisms (Burkard et al., 1990; Lasky, 1997) and in the diagnosis of certain pathologies (Tanaka et al., 1996; Jiang et al., 2000). One approach that could be interesting to try in this context, would be to separate the discrete stimuli trains into multiple bins, based on SOA, i.e., as a multivariate stimulus representation. This is similar in principle to how Di Liberto et al. (2015) represented different phonetic features, and could facilitate the study of adaptation at different SOAs.

In summary, this was a highly exploratory experiment in which we aimed to investigate the utility of TRF estimation in the derivation of canonical multiple-latency responses to high-rate click trains. While we were unable to derive canonical LAEPs—and thus canonical multiple-latency responses—due to adaptation at the higher rates—the responses that were recovered can be useful for other purposes. Future work could focus on using this approach and perhaps the suggested multivariate extension to study adaptation and look to other approaches for recovering multiple-latency responses.

Experiment 2

4.5 Introduction

In Experiment 1, we investigated the utility of TRF estimation in the derivation of canonical multiple-latency responses to high-rate click trains and determined that we should focus our efforts on other approaches. These for example could include the use of TRF estimation and continuous stimuli. As mentioned in Chapter 3, this approach has been used successfully in the past, e.g., Lalor et al. (2009) showed that using TRF estimation, it is possible to derive LAEPs to AMTs and AM BBN, Lalor and Foxe (2010) showed that it is possible to derive LAEPs to continuous natural speech, and recently, Maddox and Lee (2018) modified this approach and showed that it is possible to derive responses to continuous natural speech, from the entire auditory pathway simultaneously.

Here we aim to build upon the work of Lalor et al. (2009), Lalor and Foxe (2010), and Maddox and Lee (2018), by focusing on the derivation of multiple-latency responses to AM BBN. Specifically, we aim to demonstrate that the specificity of these responses and the efficiency of their derivation can be further improved using different stimulus representations and modelling approaches respectively. We also propose to further the investigation of their neural underpinnings through comparisons with their canonical counterparts, elicited using level-specific (LS) chirp trains (Elberling and Don, 2010; Elberling et al., 2012).

4.6 Materials and Methods

4.6.1 Subjects

13 subjects aged 23–35 years participated in this study; 5 were male. All subjects had self-reported normal hearing. The protocol for this study was approved by the Ethics Committee of the Health Sciences Faculty at Trinity College Dublin, Ireland, and all subjects gave written informed consent.

4.6.2 Stimuli

As mentioned, two different types of stimuli were used in this study, AM BBN and LS-Chirp trains.

The carrier signal for the AM BBN stimulus (Figure 4.3A) was uniform BBN with energy limited to a bandwidth of 0–24000 Hz (Figure 6.1C). Its modulating signal had a log-uniform amplitude distribution and a bottom-heavy (right-skewed) frequency (modulation rate) distribution (Figure 6.1E and G)—so chosen as it has been shown that auditory cortical areas tend to be most sensitive to AM stimuli presented at lower modulation rates (Liégeois-Chauvel et al., 2004). The modulating signal was created by first generating a discrete "template" signal, with impulse amplitudes randomly drawn from a beta distribution, i.e., $B(0.65,0.65)$, and impulse SOAs uniformly randomly set to between 0.125 s and 0.0313 s, resulting in instantaneous presentation rates of between 8 and 32 Hz. Another signal was then created by interpolating between the different impulses in the template signal, resulting in a continuous signal with a slightly more Gaussian amplitude distribution than the template—and so uniform overall given that the template signal had a slightly "U-shaped" beta amplitude distribution—and a more bottom-heavy modulation rate distribution—given that the amplitude at each impulse now only contributed to at most half of a full-cycle in the continuous signal, although typically less than that as the impulse amplitudes did not always alternate consecutively, thus, at least halving the instantaneous presentation rates to between 4 and 16 Hz. Finally, the continuous modulating signal—which was assumed to be in its SPL form—was transformed into its voltage form using:

$$x_{voltage} = 10^{((80 \times x_{SPL})/20)}$$

where $x_{voltage}$ is the modulating signal in its voltage form, x_{SPL} is the modulating signal in its SPL form, and 80 refers to the maximum presentation level in dB SPL. This was done so that when the modulating signal was applied across the transducer, it would be transformed back into its SPL form with a uniform amplitude distribution as intended.

Click stimuli have traditionally been used to elicit canonical responses, and while clicks are often considered "broadband"—due to the inherent spectral properties of impulsive stimuli—the responses they elicit in fact predominantly reflect more basal, i.e. near the base, high-frequency regions of the cochlea (Dau et al., 2000). This is due to a decrease in neural synchrony and an increase in phase cancellation as the click-induced travelling wave propagates apically along the basilar membrane (Dau et al., 2000). Chirp stimuli attempt to compensate for this by presenting their frequency components from low to high (Figure 4.3B and C), such that each component arrives at its place of maximum excitation along the basilar membrane synchronously (Elberling et al., 2007b). The LS-Chirps used in the present study,

also take the influence of stimulus level on response latency into account (Elberling et al., 2012).

Here, alternating 60 dB nHL LS-Chirp trains with uniformly distributed ±25% jitter were presented at rates of 20.1, 12.3, and 1.0 Hz (Figure 4.3B–D). These specific rates were chosen because they are suitable—in terms of having sufficiently long SOAs—for evoking canonical ABRs, MLRs, and LAEPs, respectively.

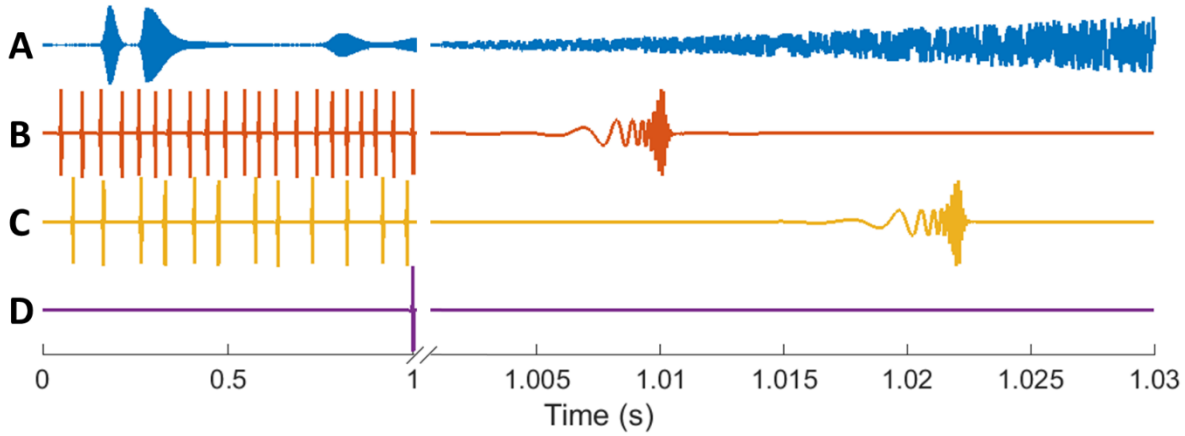

Figure 4.3: Example Segments of the Stimuli used in This Experiment.

A *– The AM BBN stimulus.* ***B–D*** *– Alternating 60 dB nHL LS-Chirp trains, with uniformly distributed +/-25% temporal jitter, presented at 20.1, 12.3, and 1.0 Hz, respectively. Please note the variable timescale on the x-axis which is there to provide both a sense of their overall time-course (left, 0 to 1 s) and their fine temporal detail (right, 1 to 1.03 s).*

4.6.3 Experimental Procedure

Subjects were presented with 80 repetitions of the same 60 s long AM BBN stimulus, 8 repetitions of the same 60 s long 20.1 Hz LS-Chirp train, 4 repetitions of the same 60 s long 12.3 Hz LS-Chirp train, and 7 repetitions of the same 60 s long 1.0 Hz LS-Chirp train, as they reclined in a comfortable chair, in a quiet, darkened room, and watched a silent animated cartoon presented on a tablet computer. The different LS-Chirp trains were repeated different numbers of times in accordance with the expected SNRs of the responses they elicit. Subjects were asked not to attend to the auditory stimuli, which were presented monaurally to their right ear. The AM BBN stimulus was presented at a peak level equivalent to that of a 1 kHz PT at 80 dB SPL, resulting in an RMS level of 59.2 dB SPL. The LS-Chirps were presented at 60 dB nHL, and calibrated using the procedure outlined by Elberling et al. (2012) and the values given in Elberling and Don (2010). All stimuli were presented using a Sound Blaster X-Fi Surround 5.1 Pro external sound card, a TPA3118D2EVM amplifier, and electromagnetically shielded

Etymotic Research ER-2 earphones, via VLC Media Player from VideoLan (http://www.videolan.org). The stimulus presentation order, i.e., for each 60 s long stimulus, was pseudorandomised to minimise any potential order effects. Compensation for the 1 ms sound tube delay introduced by the ER-2 earphones was applied post-hoc.

4.6.4 EEG Acquisition

40 channels of EEG data were recorded at 16384 Hz (analog -3 dB point of 3276.8 Hz), using a BioSemi ActiveTwo system (http://www.biosemi.com). 32 cephalic electrodes were positioned according to the standard 10-20 system. A further eight non-cephalic electrodes were also collected although only two—those over the left and right mastoids—were used in the analysis. Triggers indicating the start of each 60 s trial were encoded in a separate channel in the stimulus WAV file as three cycles of a 16 kHz tone burst. These triggers were interpreted by custom hardware before being fed into the acquisition laptop for synchronous recording along with the EEG.

4.6.5 EEG Preprocessing

The EEG data were first resampled to the appropriate rate (see below) using the *decimate* function in MATLAB (http://www.mathworks.com). The *decimate* function incorporates an 8th order low-pass Chebyshev Type I IIR anti-aliasing filter, implemented using the *filtfilt* function. A 1st order high-pass Butterworth filter was then applied with a cutoff frequency of 1 Hz, also using the *filtfilt* function. 5 Hz wide 1st order notch Butterworth filters were then applied with centre frequencies of 50, 150, 450, and 750 Hz, i.e., the electrical mains frequency, and the first three triplen harmonics, again using the *filtfilt* function. Bad channels were determined as those whose variance was either less than half or greater than twice that of the surrounding 2–4 channels, depending on location. These were then replaced through spherical spline interpolation using EEGLAB (Delorme and Makeig, 2004). Finally, the data were rereferenced to the average of the mastoids, separated into trials based on the triggers provided, and z-scored.

4.6.6 Temporal Response Function Estimation

Responses were derived using TRF estimation, and were implemented via the mTRF Toolbox (Crosse et al., 2016) using a forward modelling approach (see Chapter 3 for details). One key

consideration with these calculations is the so-called window of support, i.e., the range of time-lags over which the TRF is to be estimated. As will become clear below, this range was chosen in different ways to emphasise the different auditory latencies. In all cases, baseline correction was performed on each subject's average TRF by subtracting the mean of certain pre-stimulus values, i.e., ABR: -5 to 0 ms; MLR: -10 to 0 ms; LAEP: -20 to 0 ms, before being combined to form the grand average response.

4.6.7 AM BBN Stimulus Representations

As mentioned, an important consideration when employing TRF estimation is the choice of stimulus feature. As the defining characteristic of AM BBN is its amplitude envelope, it is the obvious feature of choice. It is worth noting that while envelopes used in previous studies have often been too slow for studying the fast dynamics of subcortical nuclei, envelopes extracted from higher sampling-rate representations of the audio signal can contain the requisite high-rate fluctuations. Here, the envelope representation was generated by first resampling the original audio signal down to 24576 Hz using the *decimate* function in MATLAB, then taking the absolute value of its Hilbert transform, before finally resampling it down to 8192 Hz.

One issue with this representation, however, is that given the strong relationship between stimulus frequency and ABR latency—due to the tonotopic nature of the cochlea—an ABR derived using a broadband envelope is likely to be temporally smeared. This could perhaps be ameliorated using an approach akin to the stacked ABR, where a number of different narrowband ABRs are first recovered, then realigned in time and summed (Don et al., 1997). In this case, such responses could be derived separately for each frequency component or perhaps simultaneously using a FB (spectrogram) representation (Ding and Simon, 2012; Di Liberto et al., 2015). However, given the temporal cost of the former and the computational cost of the latter—particularly given the high sampling rates used here—neither of these approaches were preferred.

A more efficient approach might be to first apply a compressive gammachirp auditory filter bank to the broadband audio signal (Irino and Patterson, 2006), and then use the envelope extracted from that signal to derive the ABR. This should help account for the travelling wave delay introduced by the cochlea—among other properties of the auditory periphery—without requiring multiple or multivariate analyses. The gammachirp envelope representation was generated by first resampling the original audio signal down to 24576 Hz, then bandpass filtering the audio signal into 128 logarithmically spaced frequency bands between 96 and

12288 Hz, then extracting the envelopes from each band and averaging them together, before finally resampling it down to 8192 Hz.

4.6.8 Modelling Approaches

Maddox and Lee (2018) were the first to show that it is possible to derive responses to continuous natural speech, from the entire auditory pathway simultaneously. However, the high sampling rates that are necessary for deriving the short-latency (ABR) parts of these responses make this a computationally- and memory-intensive analysis—although ameliorated in their case using a novel Fourier-based approach. This is particularly problematic as the stimulus autocovariance matrix used to derive the TRF grows quadratically with sampling rate. However, the high sampling rates that are necessary for deriving the short-latency parts of these responses are higher than are necessary for deriving the middle- and long-latency (MLR and LAEP) parts of these responses. Therefore, reducing the sampling rates to the minimum necessary when deriving each latency, should help to minimise the size of the autocovariance matrices and produce considerable improvements in modelling efficiency.

It has been suggested that the minimum sampling rate necessary when deriving ABRs to complex sounds, is 6000 Hz (Skoe and Kraus, 2010). However, we speculated that we could use even lower sampling rates without losing response fidelity. This is because we expect that TRFs mostly reflect temporal, i.e., synchronous as opposed to rate, coding of the envelope, and previous research has suggested that temporal coding in the auditory brainstem only occurs up to ~1000 Hz (Picton, 2010, 293). So, according to Nyquist, we should be able to sample as low as ~2000 Hz without a significant degradation in response morphology. Hence, a sampling rate of 2048 Hz was chosen as the minimum necessary for deriving ABRs. Sampling rates of 512 and 128 Hz were also empirically chosen as the minimum necessary for deriving MLRs and LAEPs, respectively.

The autocovariance matrix also grows quadratically with the range of time-lags. However, as before, the wide ranges that are necessary for deriving the long-latency parts of these responses, may be wider than necessary for deriving the middle- and short-latency parts of these responses. Therefore, reducing these ranges to the minimum necessary when deriving each latency, should help to minimise the size of the autocovariance matrices and produce considerable improvements in modelling efficiency. Ranges of -20 to 30 ms, -40 to 100 ms, and -75 to 425 ms, were empirically chosen as the minimum necessary for deriving ABRs,

MLRs, and LAEPs, respectively—such that any visible edge effects appeared ≥ 10 ms outside the window of interest.

Given that there are two ways of increasing modelling efficiency, i.e., by reducing the sampling rate and reducing the range, there are four ways in which these methods can be combined, i.e., using the full sampling rate full range (FSFR), reduced sampling rate full range (RSFR), full sampling rate reduced range (FSRR), and reduced sampling rate reduced range (RSRR) approaches. However, in each case a separate analysis would still be required for each latency of interest. So, to efficiently derive a multiple-latency response using just one analysis, another approach is needed.

Instead of reducing the sampling rate of the entire TRF—by reducing the sampling rate of the stimulus representation and EEG—we could vary the sampling rate of the TRF across latency by directly varying the temporal distance between adjacent time-lags in the design matrix, X, with higher sampling rates at shorter-latencies and lower sampling rates at longer-latencies. In practice, this can be achieved by replacing the usually uniformly spaced time-lag vector used to generate the design matrix—where the difference between adjacent time-lags is determined by the sampling rate of the stimulus representation and EEG—with a variably spaced time-lag vector—where the differences between adjacent time-lags are set manually.

To determine this variably spaced time-lag vector, we first started with the three empirically chosen sampling rates mentioned above—or more specifically their associated time-lag differences—to which we fit a sigmoid (Figure 4.4A). This served to give us smoother transitions from one sampling rate to the next, rather than dealing with discontinuities in the sampling rate that might produce edge effects in the analysis. This was important because, with TRF estimation, there is an inverse relationship between sampling rate and response amplitude that needs to be considered. In other words, one needs to scale the TRF as a function of sampling rate—which is now variable. Indeed, a similar issue pertains to the strength of the regularisation applied to the linear regression when calculating the TRF. The appropriate regularisation parameter, which biases the TRF estimate towards a smooth solution, varies as a function of sampling rate. So, when calculating the TRF using our proposed variable sampling rate full range (VSFR) approach, we used variable scaling and variable lambda values that were proportional to the variable sampling rate. A similar approach was taken for the pre-zero time-lags (not shown here), with sampling rates of 2048 Hz between 0 and -5 ms, 512 Hz between -5 and -20 ms, and 128 Hz between -20 and -75 ms—albeit without the subsequent fitting of a sigmoid.

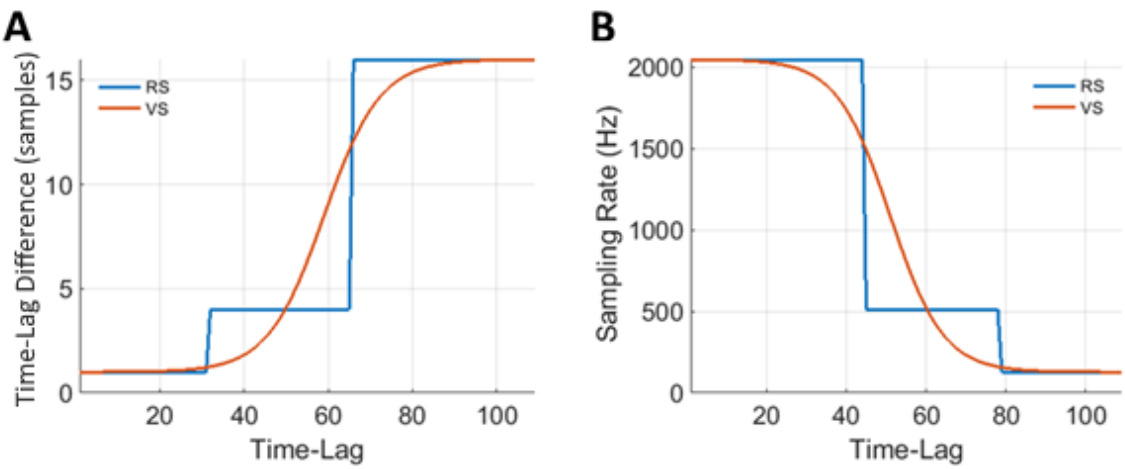

Figure 4.4: Time-lag Differences and Equivalent Reduced and Variable Sampling Rates

A *– Time-lag differences used when deriving responses as a function of time-lag, i.e. response latency.* ***B*** *– Equivalent sampling rates used when deriving responses as a function of time-lag. The blue curve shows typical values used for the reduced sampling rates (RS) of the ABR, MLR and LAEP. The red curve shows a sigmoid fit to the blue curve, representing the variable sampling rate (VS).*

4.7 Results

4.7.1 Validating our Choice of Stimulus Representation

As mentioned, the choice of stimulus representation can have a significant impact on the derived response. Specifically, we hypothesised that an ABR derived using a broadband envelope representation would likely be temporally smeared, and that this temporal smearing could be ameliorated by instead deriving the ABR using a gammachirp envelope representation. In this section, we sought to validate that decision by deriving and comparing narrowband, broadband, and gammachirp envelope ABRs. These responses were derived using the RSRR approach, and the narrowband envelopes were created by first filtering the 24576 Hz audio signal into the appropriate frequency bands, i.e., 0–1024 Hz (0–1 kHz), 1024–2048 Hz (1–2 kHz), 2048–4096 Hz (2–4 kHz), 4096–8192 Hz (4–8 kHz), and 8192–12288 Hz (8–12 kHz), using 4th order high- and low-pass Butterworth filters, implemented using the *filtfilt* function in MATLAB—which again in-effect doubles the order of the filter—and then extracting the envelopes as before.

From Figure 4.5A we can see that as expected, the latency of the main peak which we take as analogous to Wave V in the narrowband envelope ABRs, varies considerably with frequency, with responses to the high-frequency envelopes preceding those to the low-

frequency envelopes. This illustrates the degree of variation that the broadband envelope ABR must account for when finding the linear best fit, which explains the temporally broad response seen in Figure 4.5B. The gammachirp ABR on the other hand—which has had much of this variation accounted for a priori—is, as expected, larger and more sharply resolved.

It is also interesting to note from these responses that the latencies seem a little earlier than expected. For example, click ABRs are principally driven by frequency components over 4 kHz (Don et al., 1997), and so one might expect that the wave V latencies of the 4–8 kHz and 8–12 kHz envelope ABRs would coincide with those of the click ABR. That is not what we see here however with the 4–8 kHz and 8–12 kHz envelope ABRs having wave V latencies of 3.78 ms and 4.15 ms respectively, compared to the normative click ABR wave V latency of 5.63 ms (Campbell et al., 1981). However, it is difficult to compare these responses directly as the normative latencies are based on 70 dB nHL clicks while the levels of the narrowband envelopes are inherently variable. That said, the levels of the latter are lower, in which case one would expect their latencies to be later if anything. Perhaps there is simply less lag for an envelope following response than there is for an onset response.

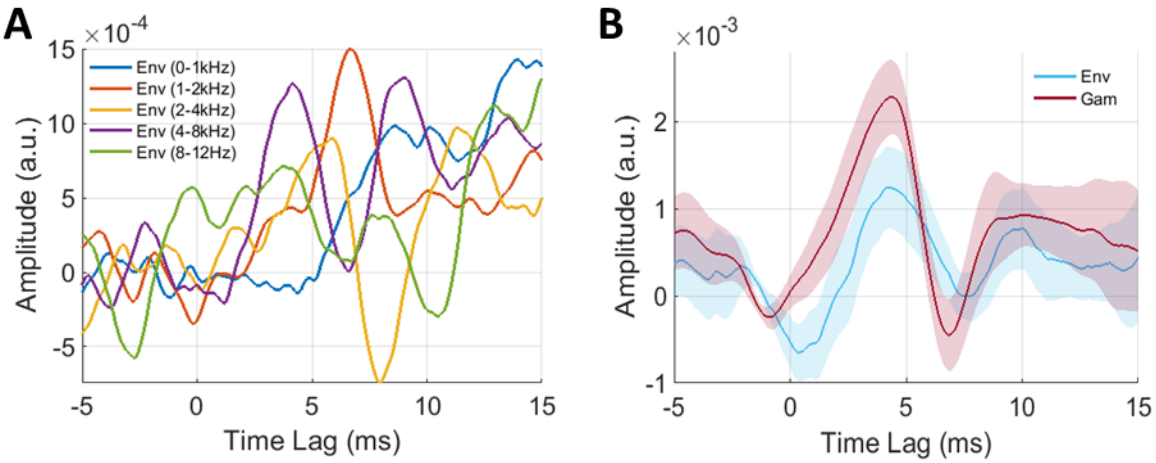

Figure 4.5: Narrowband, Broadband, and Gammachirp Envelope ABRs

A *– Narrowband envelope ABRs show considerable variation in latency as a function of frequency—a well-known phenomenon.* ***B*** *– The variation in ABR latency as a function of frequency can result in a temporally smeared ABR when using a broadband envelope (Env). Using a gammachirp envelope (Gam) representation can help produce an ABR with temporally sharper peaks. The shaded areas indicate the standard error of the mean (SEM).*

4.7.2 Evaluating the Reduced and Variable Sampling Rate Responses

In this section, we wanted to compare the computationally expensive, FSFR approach, with the RSFR, RSRR, and VSFR approaches. The purpose of these comparisons was to determine what if anything is lost with the latter approaches respective gains in efficiency, and to ensure that they still return accurate and meaningful measures of low-level auditory processing. These comparisons were carried out for each auditory latency—simply requiring the displayed range to be changed in the case of the VSFR multiple-latency response—and were focused on data recorded from the midline central electrode Cz.

Auditory Brainstem Response

Figure 4.6A shows the grand average ABR derived using the FSFR approach (8192 Hz, -75 to 425 ms). We have calculated it here as our "gold standard", as it aims to incorporate the influence of the stimulus on the response at all relevant time-lags and at sampling rates that are more than sufficiently high for all latencies. Figure 4.6A also shows two other ABRs, one that was calculated using the FSRR approach (8192 Hz, -20 to 30 ms), and one that was calculated using the RSRR approach (2048 Hz, -20 to 30 ms).

All three ABRs show a high degree of correspondence in terms of their main features i.e., relative magnitudes, latencies, and morphologies. This validates our expectation that a sampling rate of 2048 Hz would be sufficient to capture the relevant features of the ABR. However, it is noteworthy that in the time range of 9 to 15 ms, the two ABRs derived using the reduced range differ from that using the full range. This is likely because calculating the dependence of the EEG on the stimulus feature at specific time-lags, while ignoring the influence of the stimulus feature at other time-lags, leads to artificial distortions in the TRF as it tries to explain as much variance in the data as possible without access to all the dependent variables. The VSFR ABR (Figure 4.6B) on the other hand—which does take all the relevant time-lags into account—does not suffer from the same distortions. These observations were reflected in correlational analyses conducted using Pearson's correlation coefficient over the time range of 0 to 15 ms within individual subjects, i.e., the FSFR ABR was more highly correlated with the VSFR ABR than either of the other two ABRs (see Table 4.1).

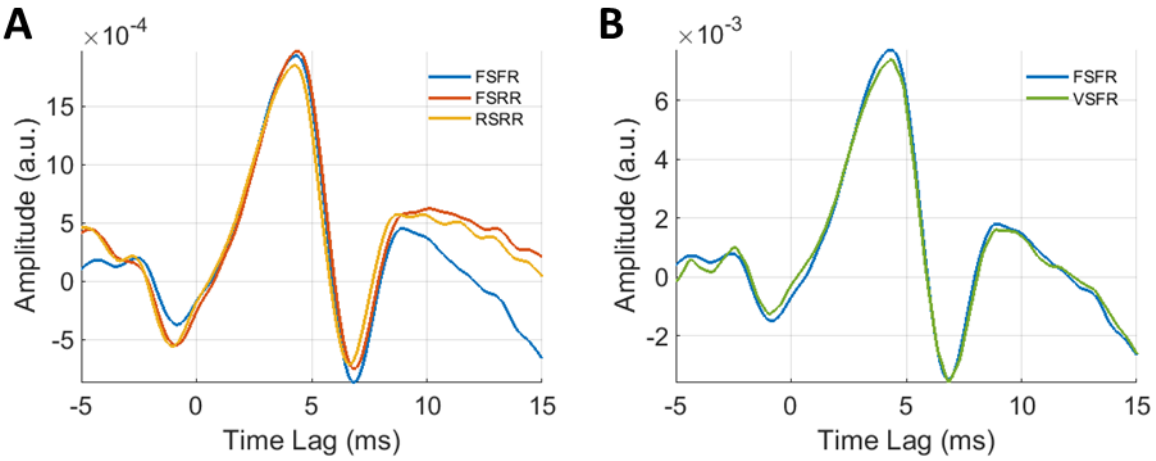

Figure 4.6: ABRs Derived using the FSFR, FSRR, RSRR, and VSFR Approaches

A *– ABRs derived using the full sampling rate full range (8192Hz, -75 to 425 ms; FSFR), full sampling rate reduced range (8192 Hz, -20 to 30 ms; FSRR), and reduced sampling rate reduced range (2048 Hz, -20 to 30 ms; RSRR) approaches. While displaying a high degree of correspondence, differences in the responses are apparent—especially from 9 to 5 ms—when using the reduced range approaches.* ***B*** *– ABRs derived using the FSFR and variable sampling rate full range (VSFR) approaches. The VSFR ABR recapitulates the features of the FSFR ABR.*

Table 4.1: Comparison Between ABRs derived using the FSFR and Other Modelling Approaches

Modelling Approach	Mean r	SD	p
FSRR	0.96	0.05	< 0.001
RSRR	0.94	0.04	< 0.001
VSFR	0.99	0.01	< 0.001

Middle-Latency Response

Figure 4.7A shows the grand average MLRs derived using the FSFR (8192 Hz, -75 to 425 ms), FSRR (8192 Hz, -40 to 100 ms), and RSRR (512 Hz, -40 to 100 ms) approaches. All three MLRs show a high degree of correspondence in terms of their main features. However, in the time range of 60 to 80 ms, the two MLRs derived using a reduced range differ from that using the full range. This is not unexpected, however, given that a similar effect was seen with the reduced range ABRs. It is also interesting to note that all three MLRs display a “N1” peak at ~60 ms, which is much earlier than would be expected from a canonical response.

Similar to the VSFR ABR, the early peaks of the VSFR MLR are better resolved than in the FSFR MLR, even though the latter was derived using a higher sampling rate (Figure 4.7B). While the classic MLR peaks, i.e., Na, Pa, and Nb, and even early LAEP peaks, i.e.,

"P1" and N1, seem to be present in both responses, the latencies of the Pa, Nb, and P1 peaks seem earlier for the VSFR response, and the N1 peak seems later. These observations were reflected in correlational analyses conducted over the time range of 0 to 80 ms within individual subjects, i.e., the FSFR MLR was more highly correlated with the FSRR MLR and RSRR MLR than the VSFR MLR (see Table 4.2).

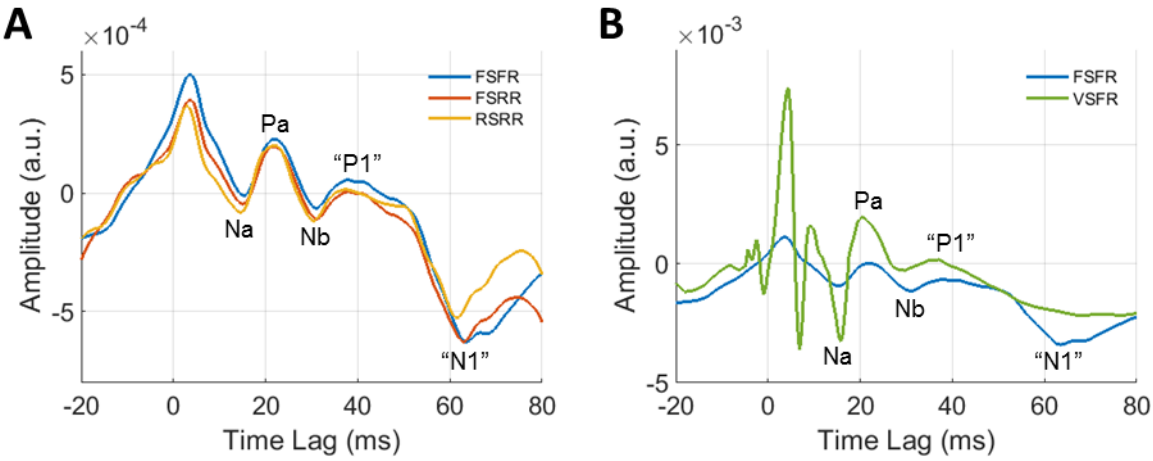

Figure 4.7: MLRs Derived using the FSFR, FSRR, RSRR, and VSFR Approaches

A *– MLRs derived using the FSFR (8192Hz, -75 to 425 ms), FSRR (8192 Hz, -40 to 100 ms), and RSRR (512 Hz, -40 to 100 ms; RSRR) approaches. While displaying a high degree of correspondence, differences in the responses are apparent—especially from 60 to 80 ms—when using the reduced range approaches.* ***B*** *– MLRs derived using the FSFR and VSFR approaches. The VSFR MLR displays similar peaks to the FSFR MLR albeit at somewhat different latencies. Please note, the FSFR MLR peaks were labelled in both figures.*

Table 4.2: Comparison Between MLRs derived using the FSFR and Other Modelling Approaches

Modelling Approach	Mean r	SD	p
FSRR	0.97	0.05	< 0.001
RSRR	0.95	0.03	< 0.001
VSFR	0.63	0.16	< 0.001

<u>Late Auditory Evoked Potential</u>

Figure 4.8A shows the grand average LAEPs derived using the FSFR (8192 Hz, -75 to 425 ms) and RSFR (128 Hz and -75 to 425 ms) approaches. The FSRR and RSRR approaches could not be used here, as with the LAEP, the range cannot be reduced any further. Both LAEPs show a high degree of correspondence in terms of their main features. Interestingly, both responses display a "N1" peak at ~79 ms rather than at ~60 ms as was seen in the MLRs. As was the case for both the VSFR ABR and MLR, the VSFR LAEP displays stronger early peaks

than the FSFR LAEP. While the P1 peak is obscured in the FSFR response, the N1, P2, and N2 peaks seem to correspond relatively well between the two responses. These observations were reflected in correlational analyses conducted over the time range of 0 to 400 ms within individual subjects, i.e., the FSFR LAEP was more highly correlated with the RSFR LAEP than the VSFR LAEP (see Table 4.3).

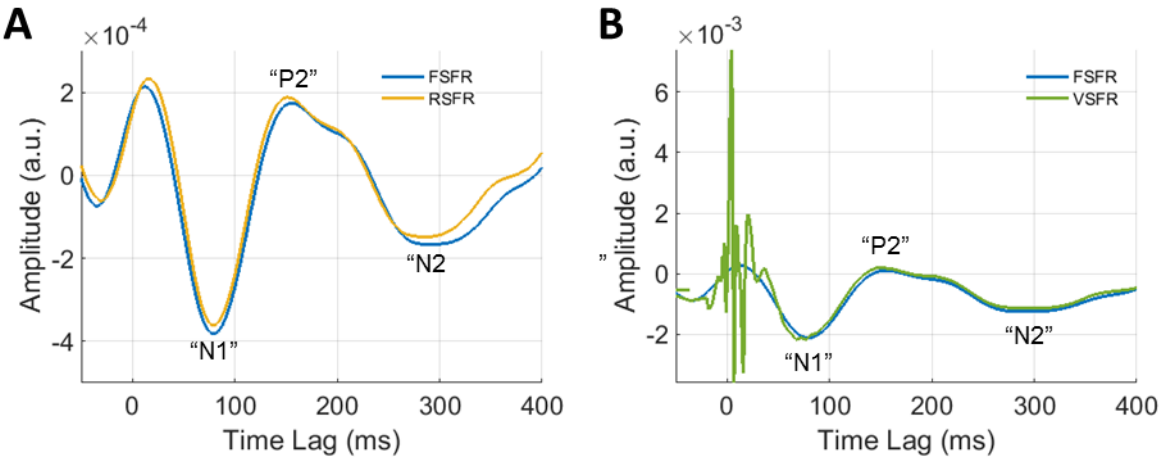

Figure 4.8: LAEPs Derived using the FSFR, RSFR, and VSFR Approaches

A *– LAEPs derived using the FSFR (8192Hz, -75 to 425 ms), and RSFR (128 Hz, -75 to 425 ms) approaches. The RSFR approach was used as the range could not be reduced any further, and because of this, no major disparities were seen at the response edges.* ***B*** *– LAEPs derived using the FSFR and VSFR approaches. Both responses show a high degree of correspondence. Please note, the FSFR LAEP peaks were labelled in both figures.*

Table 4.3: Comparison Between LAEPs derived using the FSFR and Other Modelling Approaches

Modelling Approach	Mean r	SD	p
RSFR	0.98	0.01	< 0.001
VSFR	0.65	0.13	< 0.001

4.7.3 Comparing the Variable Sampling Rate Full-Range Multiple-Latency Response to Canonical Responses

Having established that the VSFR multiple-latency response returns what seems like reasonable measures of processing across the auditory hierarchy, we wished to characterise it further through comparisons with canonical responses. These canonical responses were elicited using 60 dB nHL LS-Chirp trains and were derived using the RSFR approach. Comparisons were performed for each auditory latency—again simply requiring the displayed range to be changed in the case of the VSFR multiple-latency response—and were focused on data

recorded from the midline central electrode Cz. To facilitate these comparisons, the chirp responses were scaled such that the amplitude of the first prominent peak for a given latency, i.e., wave V, Na, and N1, for the ABR, MLR, and LAEP respectively, matched those of the VSFR multiple-latency response.

Auditory Brainstem Response

Figure 4.9A shows the grand average VSFR ABR as well as the grand average 20.1 Hz 60 dB nHL LS-Chirp train ABR. While the overall morphologies of the two responses were similar, the waves after the VSFR wave V appeared to be slightly more resolved, although we do not wish to overstate this given that the two responses were elicited using different stimuli. The peak of the VSFR wave V appears slightly later at this electrode, although a measure of the ABR over the entire scalp known as the global field power (GFP; Lehmann, 1987), as well as inspection of the other channels (not shown here) suggests that both responses actually peaked at the same time (~3.42 ms; Figure 4.9B). While the latencies of both responses might appear early, it is known that LS-Chirp ABRs typically display peaks ~1.5 ms earlier than those of click ABRs at lower levels in normal-hearing subjects (Elberling et al., 2012).

These observations were not strongly reflected in correlational analyses conducted over the time range of 0 to 15 ms within individual subjects (mean $r = 0.55$, $SD = 0.39$, not all $p < 0.05$), although more so across group averages ($r = 0.85$, $p < 0.001$), and in their topographic distributions (Figure 4.9C), with both responses displaying a left lateralised unimodal distribution consistent with a brainstem response elicited by a stimulus presented to the right ear. Overall, these findings suggest that, indeed, the VSFR ABR is reliably reflecting brainstem activity.

Middle-Latency Response

Figure 4.10A shows the grand average VSFR MLR as well as the grand average 12.3 Hz 60 dB nHL LS-Chirp train MLR. While the overall morphology of the two responses was similar, there were some differences in the latencies of their respective peaks. While the Na peaks of both responses seem to co-occur, the VSFR Pa, Nb, and “P1” peaks appear much earlier, and the “N1” peak appears much later at this electrode.

The similarities in overall morphology in the grand average data were not reflected strongly in correlational analyses conducted over the time range of 0 to 80 ms within individual

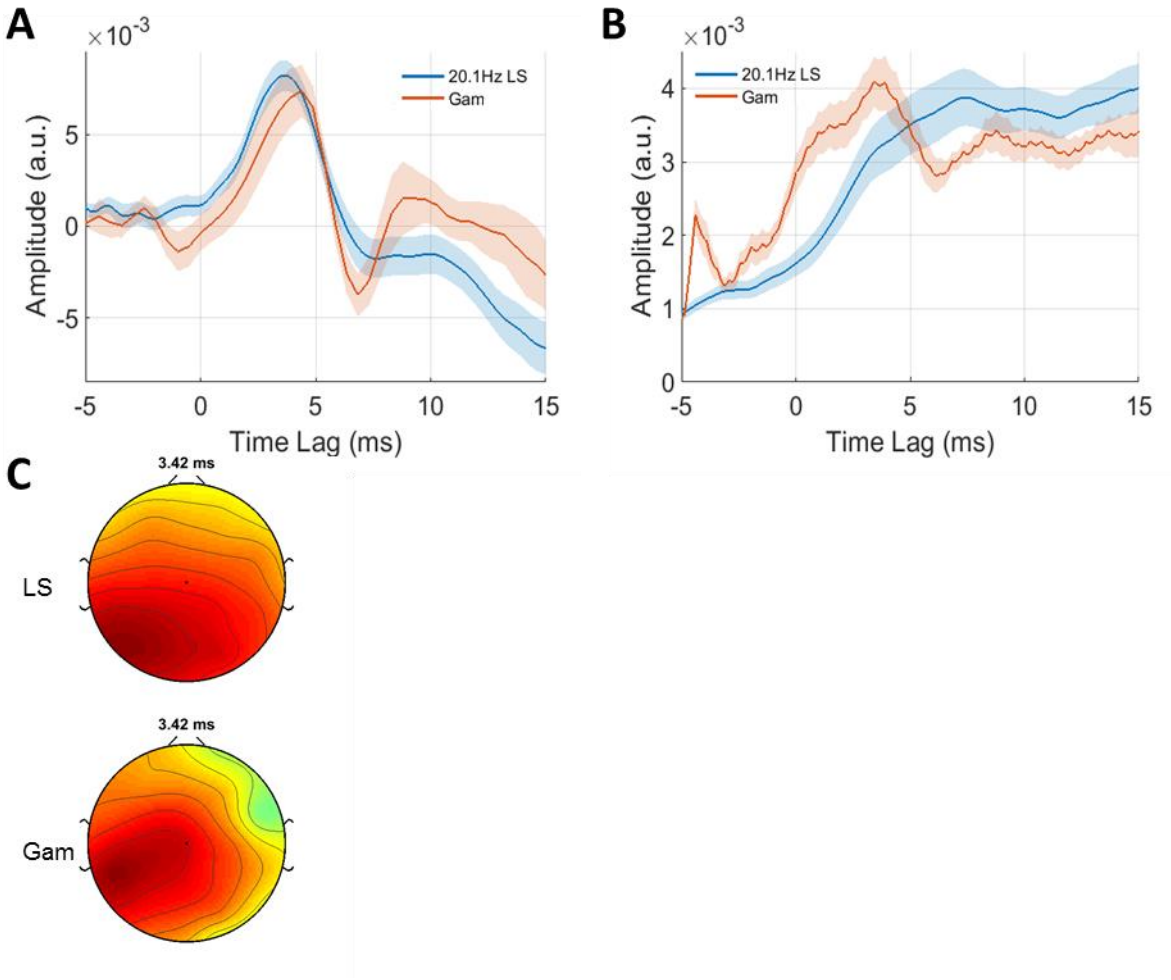

Figure 4.9: Comparison of VSFR and 20.1 Hz LS-Chirp train ABRs.

A *– Both waveforms show a clear wave V, although the VSFR ABR shows some sharper later waves.* ***B*** *– A measure of the two ABRs across the entire scalp, known as global field power (GFP). This measure is temporally sharper for the VSFR ABR.* ***C*** *– Topographic distributions of the ABRs at the estimated wave V latency. The topographies are similar likely indicating common neural generators. The shaded areas indicate the SEM.*

subjects (mean r = 0.03, SD = 0.31, not all $p < 0.05$), although more so across group averages ($r = 0.51$, $p < 0.001$), and also in their topographic distributions (Figure 4.10C).

Late Auditory Evoked Potential

Figure 4.11A shows the grand average LAEP as well as the grand average 1.0 Hz 60 dB nHL LS-Chirp train LAEP. While the overall morphologies of the two responses were similar, there were some differences in the latencies of their respective peaks, e.g., the VSFR LAEP N1 and P2 peaks appear earlier, and the N2 peak somewhat later at this electrode.

These observations were again not reflected strongly in correlational analyses conducted over the time range of 0 to 400 ms within individual subjects (mean r = 0.19, SD = 0.30, not all $p < 0.05$), although more so across group averages ($r = 0.44$, $p < 0.001$), and also in their topographic distributions (Figure 4.11C). It is also worth noting that—in-line with the findings of Maddox and Lee (2018)—the relative amplitudes between the different latencies in the VSFR LAEP are quite different to what is seen in the canonical responses, i.e., the magnitudes at later latencies are much smaller than expected. Overall, these findings suggest that, indeed, the VSFR LAEP is reflecting cortical activity.

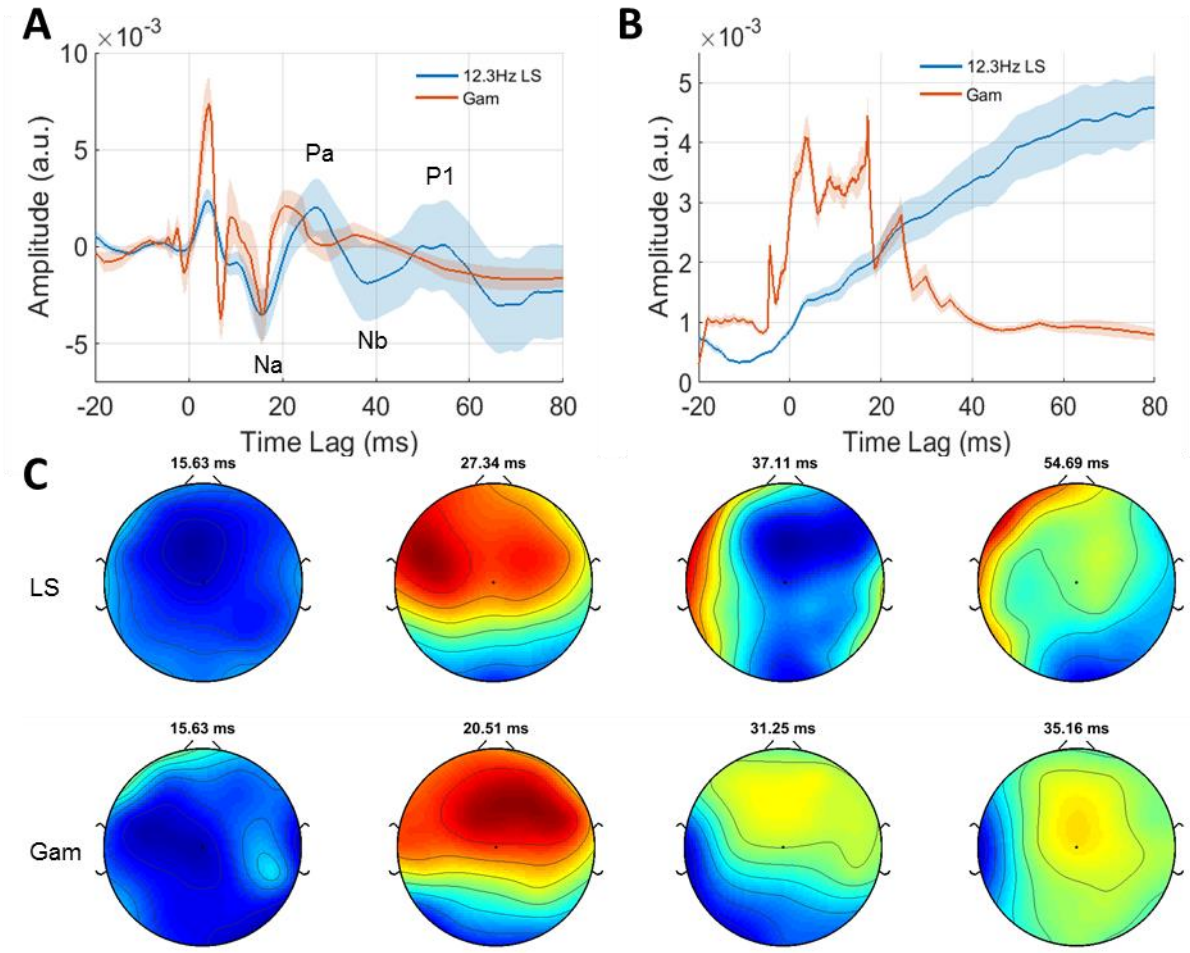

Figure 4.10: Comparison of VSFR and 12.3 Hz LS-Chirp train MLRs.

A *– Both waveforms display similar peaks albeit with somewhat different latencies. Please note, the chirp MLR peaks were labelled in this figure.* ***B*** *– GFPs of both MLRs.* ***C*** *– Topographic distributions of the MLRs at their respective estimated Na, Pa, Nb, and P1 latencies. The topographies for certain peaks are similar likely indicating common neural generators. The shaded areas indicate the SEM.*

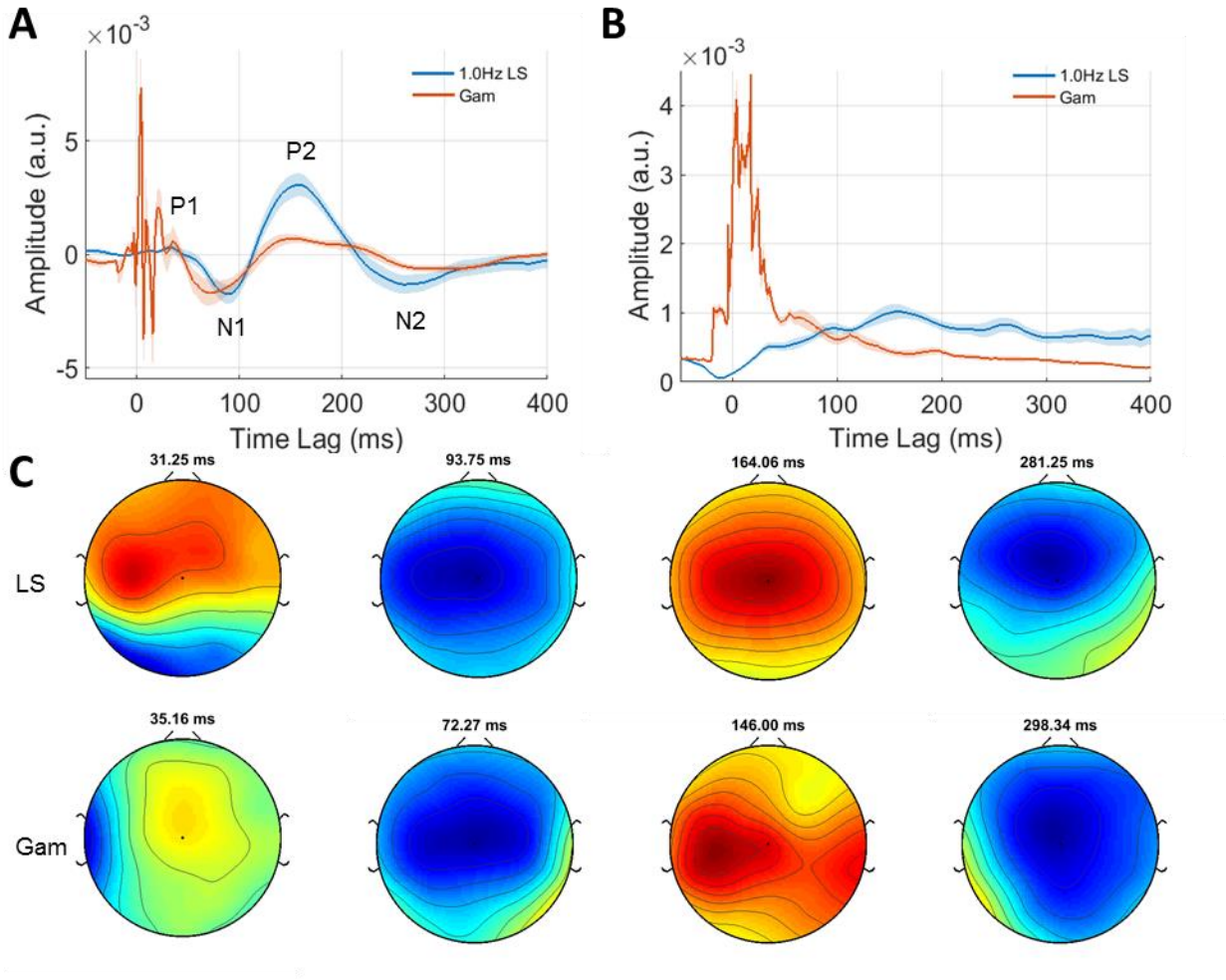

Figure 4.11: Comparison of VSFR and 1.0 Hz LS-Chirp train LAEPs.

A *– Both waveforms display similar peaks albeit with somewhat different latencies. Please note, the chirp LAEP peaks were labelled in this figure.* ***B*** *– GFPs of both LAEPs.* ***C*** *– Topographic distributions of the LAEPs at their respective estimated P1, N1, P2, and N2 latencies. The topographies for certain peaks are similar likely indicating common neural generators. The shaded areas indicate the SEM.*

4.8 Discussion

The high sampling rates that are sometimes necessary when deriving TRFs can pose several challenges. Here we have shown that by choosing a suitable stimulus representation and allowing the sampling rate of the TRF to vary across latency—with appropriate scaling and regularisation—we can efficiently derive robust multiple-latency responses from the human auditory system.

Specifically, we have shown the envelope to be a useful representation, as it can retain contributions from higher-frequency components, even when downsampled. The gammachirp

envelope representation, in particular, proved useful for ameliorating the temporal smearing introduced by the travelling wave delay of the cochlea. One issue with the gammachirp envelope representation, however, is that the parameters used to calculate it are predetermined, i.e., based on normative data. Subject-specific compensations similar to the stacked ABR approach (Don et al., 1997) would perhaps be preferable. This would require narrowband responses to be derived either separately—as before except maybe using the full range—or simultaneously using a frequency-binned representation (Ding and Simon, 2012; Di Liberto et al., 2015). These kinds of analyses would have been quite onerous to conduct before but should be facilitated by the current approach. We have also shown that because we expect the TRF to mostly reflect temporal coding of the envelope, much lower sampling rates can be used than previously thought. This might allow researchers to bypass the hardware limitations usually associated with higher sampling rates, and garner greater spatial resolution using denser arrays of electrodes.

One factor that will be critical when using the VSFR approach in future, is the tuning of the various parameters involved. In the present study, there were several decisions that needed to be made in terms of the analysis. For example, we elected to vary the sampling rate according to a sigmoid (Figure 4.4). This seems to have worked well in terms of balancing efficiency and response fidelity, but it may be that other mappings would have given better control over this trade-off, e.g., two conjoined sigmoids, one to transition from the ABR to MLR, and another to transition from the MLR to LAEP. Furthermore, this initial choice then constrained us to consider similar mappings for the scaling and regularisation of the responses. Again, deriving TRFs with different sampling rates produce responses at different scales that need to be regularised to different degrees. Balancing this scaling and regularisation required substantial care in the present work.

To illustrate some of the decisions that had to be made, take our FSFR, FSRR, and RSRR MLRs for example (Figure 4.7A). All three MLRs display a clear “N1” at ~60 ms, similar to the one reported by Maddox and Lee (2018). However, in Figure 4.8A, when using (lower resolution) LAEP parameters, both responses display a clear “N1” at ~79 ms. While it might be easy to conclude that the resolution of the latter is not high enough and that the peak is becoming obscured, it could also be argued that the resolution of the former is too high and that the TRF is overfitting. This was certainly the case just after the N1 in the former but does not seem to have been the case for Maddox and Lee (2018). Similar challenges were experienced with the “P1”. While it is beneficial to have such a high degree of control over the

response derivation, further work will be needed to elucidate the neural underpinnings of these peaks—perhaps using complementary recording methods with greater spatial resolution, such as magnetoencephalography (MEG)—and to home in on the optimal parameters for deriving them. That was beyond the scope of the current study, however, so parameters were empirically chosen to provide as high a resolution as possible, without overfitting.

It is also interesting to note that overall, the continuous responses are smaller and earlier than the discrete canonical responses. Maddox and Lee (2018) found this also and suggested that the reduced amplitudes could possibly be due to adaptation, and that these adaptation effects would be greater for the later latencies. With regards to the continuous responses being earlier, perhaps there is simply less lag between the stimulus and response for envelope-following responses than there are for onset responses. Further work will be required to test this hypothesis.

One concern with the VSFR approach is that we are effectively downsampling the TRF without the use of an anti-aliasing filter. While this is technically true, both the envelope and EEG are dominated by lower frequency components and so if any aliasing does occur it is likely to be minimal, i.e., the values chosen are likely to be close to the value of the underlying low-frequency component. Furthermore, the equivalent instantaneous sampling rates used at each lag are sufficiently high to capture the dominant frequency components in each latency. It is for these reasons, therefore, that aliasing is not deemed to be of great concern with the VSFR approach.

While 80 s of data were used for the AM BBN analyses, less data would likely have sufficed. As with TDA, the inclusion of more and more data in a TRF analysis tends to come with diminishing returns. However, as this was an exploratory study, we wanted to ensure that we used enough data to test our hypothesis, and not leave its validation potentially inconclusive on account of insufficient SNR. Now that we have shown that this approach works, future work could focus on determining the minimum amount of data, number of channels, etc., that would be needed to efficiently conduct an analysis of this type—a crucial step if such methods are to be used in clinical environments.

In summary, we have introduced a novel approach for efficiently deriving multiple-latency responses from the human auditory system. The interpretability of these responses, combined with the flexibility and efficiency with which they can be derived, should make this an attractive approach to researchers interested in studying the hierarchical processing of complex and natural sounds. Such work could lead to more diagnostically useful objective

measures of hearing function along the auditory pathway, that could help to better elucidate the source of any dysfunction, while permitting a wider variety of stimuli to be used. Future work could focus on optimising the tuning of the various parameters, investigating the neural underpinnings of the various waves, and testing this approach with continuous natural speech.

Chapter 5. MAMTA: Multiplexed Amplitude Modulated Tone Audiometry

5.1 Introduction

In this chapter, a novel TRF estimation approach for objectively determining hearing thresholds using multiplexed, i.e., multiple, mixed, AMTs is presented. Considerations of stimulus type, stimulus representation, and modelling approach are discussed, and the performance of this approach evaluated through comparisons with thresholds recovered using PTA, in both normal hearing and hearing loss populations.

In Chapter 4, we built on the work of Lalor et al. (2009), Lalor and Foxe (2010), and Maddox and Lee (2018), and proposed a complementary TRF estimation approach for efficiently deriving multiple-latency responses to AM BBN. As part of that work, we also showed that it was possible to derive frequency-specific ABRs to AM BBN using TRF estimation and narrowband envelope representations extracted from the AM BBN stimulus. While this was not entirely unexpected given that frequency-specific responses to AMTs have been derived before (Lalor et al., 2009), and spectrogram representations have been successfully used in the past (Ding and Simon, 2012; Di Liberto et al., 2015), it does raise an interesting question around the use of TRF estimation in objective audiometry.

As mentioned in Chapter 1, audiometric profiles are often established using PTA. With PTA, a subject is typically presented with 1–3 s excerpts of a PT at a fixed level and asked if they can hear it. The level is then adjusted in descending and ascending runs until the lowest level at which the subject can hear that PT, i.e., their hearing threshold for that PT, has been determined. This process is then repeated for the other PTs being tested, until a full audiometric profile has been established (BSA, 2011). While this approach has been instrumental in the

assessment of hearing function for decades, it is limited in terms of its diagnostic abilities (Musiek et al., 1994; Ruggles et al., 2011), and is not suitable for use with young children or those with a diminished capacity to respond (Downs et al., 1966).

Audiometric profiles can also be established objectively using AEPs, i.e., using ABRs (Davis and Hirsh, 1979; Don et al., 1979; Picton et al., 1979), MLRs (Musiek and Geurkink, 1981), or LAEPs (Tyberghein and Forrez, 1971). In this case, a subject is typically presented with brief, repeated, frequency-specific stimuli, such as clicks in notched-noise or tone-bursts, at a fixed level, and the response to each stimulus added to a running average. The SNR of this running average typically increases over time as more and more responses are added, and as soon as a decision has been made that a response is present at a given level, the level is reduced, and the stimulus presentation and response evaluation begin again. These steps are repeated until the lowest level at which a response can reliably be detected, i.e., the subject's objective threshold for that stimulus, has been determined. The entire process is then repeated for the other frequencies being tested, until the tester is satisfied that a reasonable audiometric profile has been established.

Unlike PTA, this approach can be used with young children, and can provide evidence of auditory system dysfunction not detectable with PTA (Musiek et al., 1994). One shortcoming of this approach, however, is that response recovery can be relatively slow, as only one stimulus can be presented—and thus one frequency and level assessed—at a time. This can be improved by interweaving stimuli of different frequencies (Ross et al., 1999), presenting more complex stimuli, e.g., multitone complexes (Bardy et al., 2015), or recording from both ears simultaneously—provided that the presentations are alternated and that it is the ABR that is being recovered (Picton, 2010, p. 175). While other measures such as the ASSR permit multiple continuous narrowband stimuli to be presented simultaneously—at different rates (Lins and Picton, 1995)—this comes at the cost of any temporal resolution in the response, i.e., it is typically analysed in the frequency domain.

In this study we aim to describe a novel approach for objectively determining hearing thresholds, using TRF estimation and multiplexed AMTs (MAMTs). Unlike other approaches where stimuli are typically presented at one level at a time, here the level of each AMT is varied continuously, and the function at each level assessed post-hoc. Also, instead of analysing data on a channel-by-channel basis, backward modelling is employed in an effort to exploit all the available data in a multivariate way (see Chapter 3 for details). It was hoped that this would provide a more sensitive measure of auditory processing, having already proven useful in other

challenging multiple-stimuli environments, e.g., the cocktail party environment (O'Sullivan et al., 2014). The relative performance of this approach will be evaluated through comparison with thresholds determined using PTA, in both normal hearing and hearing loss populations.

5.2 Materials and Methods

5.2.1 Subjects

18 subjects aged 24–66 years participated in this study; 10 were male. 9 of these subjects had normal hearing, i.e., PT thresholds of ≤ 20 dB HL, at all octave frequencies between 250–8000 Hz, in the ear being tested. The 9 remaining subjects had varying degrees of hearing loss, i.e., PT thresholds of > 20 dB HL, at one or more octave frequencies between 250–8000 Hz in the ear being tested. Hearing thresholds were determined through PTA performed by a qualified audiologist. The protocol for this study was approved by the Tallaght Hospital / St. James's Hospital Joint Research Ethics Committee. All subjects gave written informed consent and were compensated for their time.

5.2.2 Stimuli

Six PT carriers at the standard audiometric octave frequencies of 250–8000 Hz were amplitude modulated and multiplexed before being presented as one complex sound. Apart from the fact that these are the frequencies typically tested during PTA, the fact that they were each placed one octave apart also helped to minimise the likelihood of any unwanted inter-frequency interactions.

Great care needed to be taken when designing these modulating signals. Specifically, it was critical that they provided enough stimulation, both supra- and subthreshold, for a clear distinction at threshold to be made. The "ideal" amplitude distributions for these modulating signals, therefore, are likely ones which are heavily weighted around the subject's individual thresholds—as then most time would be spent stimulating the levels most relevant to threshold determination. As these would not be known a priori, however—or at least would not be in a clinical setting—log-uniform amplitude distributions were chosen in compromise. Bottom-heavy (right-skewed) frequency (modulation rate) distributions were also chosen.

It was also important to ensure that any differences between the individual AMT responses were due to frequency-specific neurological differences rather than any differences

between the stimuli themselves. To this end, one modulating signal was first created, by generating signals with discrete amplitude values with the desired statistical properties, and then interpolating between those discrete points to provide a smooth transition from one modulation amplitude to the next. This modulating signal was then circularly shifted by 10 s, five times, to technically create six different modulating signals, each having the exact same spectrotemporal properties. If the Pearson's correlation coefficient between any two of these modulating signals was > 0.01, a new initial modulating signal was generated, and the entire process restarted, otherwise these six modulating signals were used to amplitude modulate the PT carriers.

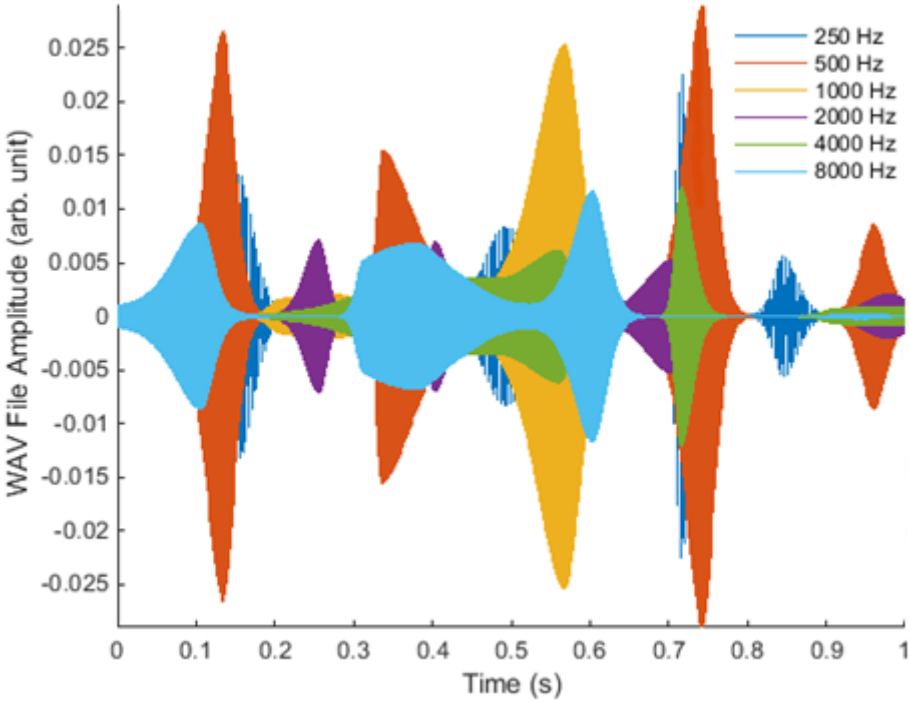

Figure 5.1: Example Segments of the Calibrated AMTs Used in This Study

250, 500, 1000, 2000, 4000, and 8000 Hz AMTs, calibrated to compensate for the frequency response of the custom hardware setup used to present them.

5.2.3 Experimental Procedure

Subjects were presented with 100 repetitions of the same 60 s long MAMT stimulus, as they reclined in a comfortable chair, in a darkened acoustically and electrically isolated room, and watched a silent animated cartoon or film presented on a tablet computer. Subjects were asked not to attend to the auditory stimuli, which were presented monaurally to either their right ear—for the normal hearing subjects—or to whichever ear had the most appropriate audiometric profile for this study—for the hearing loss subjects. Each individual AMT stimulus was

presented at a peak level of 80 dB SPL. When these were combined to form the MAMT stimulus, it theoretically would have had a peak level of

$$L_{\max(MAMT)} = L_{\max(AMT)} + 20\log_{10} n = 80 + 20\log_{10} 6 = 95.6\ dB\ SPL$$

assuming that all six AMT stimuli happened to peak at the same time. The stimuli were presented using a Sound Blaster X-Fi Surround 5.1 Pro external sound card, a TPA3118D2EVM amplifier, and electromagnetically shielded Etymotic Research ER-2 earphones, via VLC Media Player from VideoLan (http://www.videolan.org). The 70+ dB interaural isolation provided by the ER-2 headphones reduced the need for masking. Compensation for the 1 ms sound tube delay introduced by the ER-2 earphones was applied post-hoc.

5.2.4 EEG Acquisition

40 channels of EEG data were recorded at 16384 Hz (analog -3 dB point of 3276.8 Hz), using a BioSemi ActiveTwo system (http://www.biosemi.com). 32 cephalic electrodes were positioned according to the standard 10-20 system. A further eight non-cephalic electrodes were also collected although only two—those over the left and right mastoids—were used in the analysis. Triggers indicating the start of each 60 s trial were encoded in a separate channel in the stimulus WAV file as three cycles of a 16 kHz tone burst. These triggers were interpreted by custom hardware before being fed into the acquisition laptop for synchronous recording along with the EEG.

5.2.5 EEG Preprocessing

Despite being recorded at 16384 Hz, the EEG data were first resampled to 128 Hz to facilitate the development and testing of the proposed approach. This was done using the *decimate* function in MATLAB (http://www.mathworks.com) and implemented using the *filtfilt* function. Next custom high- and low-pass filters were applied, with roll-offs of 24 dB/octave and cutoff frequencies of 1 and 30 Hz respectively, also using the *filtfilt* function. Bad channels were determined as those whose variance was either less than half or greater than twice that of the surrounding 2–4 channels depending on location. These were then replaced through spherical spline interpolation using EEGLAB (Delorme and Makeig, 2004). The data were then rereferenced to the average of the mastoids, separated into trials based on the triggers provided

and z-scored. Finally, a denoising technique known as joint decorrelation (de Cheveigné and Parra, 2014) was used to enhance the reproducibility of responses across trials.

5.2.6 Temporal Response Function Estimation

Responses were derived using TRF estimation, and were implemented via the mTRF Toolbox (Crosse et al., 2016) using a backward modelling approach (see Chapter 3 for details). The use of reconstruction accuracy, i.e., the accuracy with which the stimulus representation can be reconstructed by the model, in the current study is based on the hypothesis that there is a relationship between reconstruction accuracy and hearing loss. To elaborate, if the stimulus was heard by a normal hearing subject, one might expect almost all of the stimulus feature to have been encoded in the brain, resulting in a high reconstruction accuracy, but if the stimulus was heard by a hearing loss subject, one might expect much less of the stimulus feature to have been encoded in the brain, i.e., just the suprathreshold part, resulting in a lower reconstruction accuracy. This issue is obviously more nuanced than this, but if broadly true, one can see the potential benefits of taking such an approach.

5.2.7 Stimulus Representation

The choice of stimulus feature and the way in which it is represented can have a significant impact on the derived response. As the defining characteristic of the AMTs are their amplitude envelopes (modulating signals), it was the obvious feature of choice.

5.2.8 Comparative Analyses

In PTA, PT excerpts are typically presented in dB HL, meaning that their levels have been offset by various amounts to account for normative PT thresholds, i.e., typical PT thresholds for normal hearing subjects. For example, if normal hearing subjects are typically unable to hear a 1 kHz PT until it reaches 5.5 dB SPL (ISO 389-2, 1994), then a 1 kHz PT would be offset by +5.5 dB SPL when presented in dB HL, i.e., 0 dB HL = 5.5 dB SPL. To establish normative PT thresholds, enough individual PT thresholds need to first be determined in dB SPL using PTA. This type of behavioural threshold determination is not possible with AMTs, however, as the sound level is continuously varying, and so any subjective determination of threshold in this manner would likely only reflect the threshold of the largest peak in the AMT.

It would seem then, that the only way to determine normative AMT thresholds might be to do so objectively.

This poses a problem when it comes to validating the current approach, as ideally, we would compare the AMT reconstruction accuracies with behaviourally determined AMT thresholds, which we cannot, and to do so with objectively determined AMT thresholds would be circular. Instead, it was decided that the AMT reconstruction accuracies would be compared with PT thresholds (in dB SPL), which while likely not equivalent—given the former's continuously varying nature—are likely correlated. This should not be hugely consequential when looking at the overall relationship between reconstruction accuracy and hearing loss—as it should mainly affect the relative offsets—but will be important to consider when validating the objectively determined AMT thresholds.

The global relationship between reconstruction accuracy and hearing loss was assessed by examining the slopes of regression lines fit to the data using the *polyfit* function in MATLAB and evaluated using the *polyval* function.

5.3 Results

5.3.1 Examining the Relationship between Reconstruction Accuracy and Hearing Loss

Uncorrected

Before attempting to determine individual thresholds, it seemed pragmatic to test our hypothesis, that there is a relationship between reconstruction accuracy and hearing loss at the group level. To this end, we recovered reconstruction accuracies for each AMT, for each subject, and plotted them against frequency and PT threshold (Figure 5.2).

As can be seen in Figure 5.2, there does seem to be a relationship between reconstruction accuracy and hearing loss at certain frequencies, i.e., the greater the hearing loss the lower the reconstruction accuracy—and vice versa—although this relationship seems to be reversed at 500 and 1000 Hz. These observations were reflected in the slopes of regression lines fit to the data at each frequency (Figure 5.3), with all but the 500 and 1000 Hz regression lines having negative slopes (see Table 4.3). That said, this relationship is likely confounded by individual differences in reconstruction accuracy, i.e., the fact that two subjects with the same degree of hearing loss can have different reconstruction accuracies.

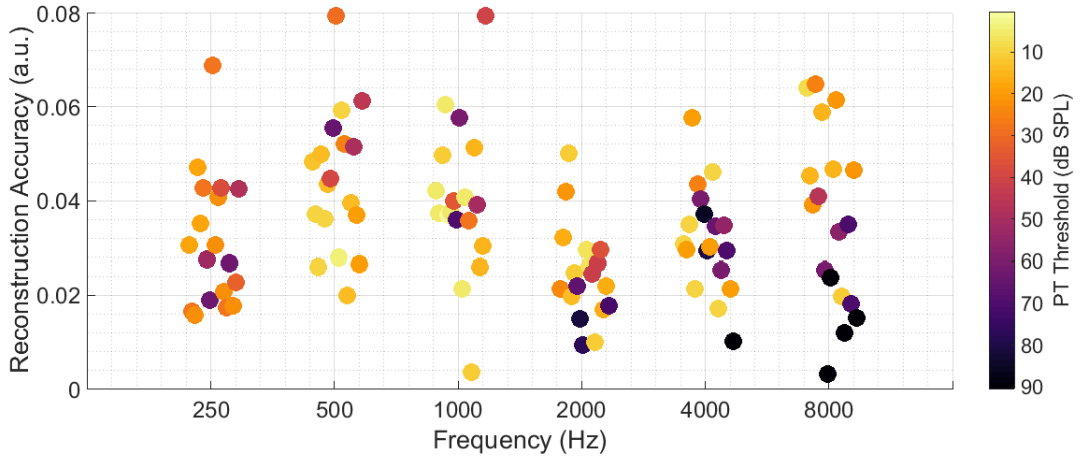

Figure 5.2: Uncompensated Global Relationship between Reconstruction Accuracy and Hearing Loss.

Reconstruction accuracies for each AMT, for each subject, plotted against frequency and PT threshold. While there does seem to be a relationship between reconstruction accuracy and hearing loss at certain frequencies, this relationship is likely confounded by individual differences in reconstruction accuracy.

<u>Corrected</u>

To compensate for these individual differences in reconstruction accuracy, subject-specific correction factors were calculated and applied. Specifically, self-normalised correction factors were calculated to map each subject's 250 Hz reconstruction accuracy onto the 250 Hz regression line. This was done to ensure that two subjects with the same degree of hearing loss would have the same reconstruction accuracy at that frequency. The data at 250 Hz was chosen as that was the frequency at which subjects had the most similar degree of hearing loss—in terms of standard deviation (STD = 14.8 dB) as determined using the *std* function in MATLAB. These subject-specific correction factors were then applied to the corresponding reconstruction accuracies at each of the other frequencies, and the resulting data plotted against frequency and PT threshold (Figure 5.4).

As can be seen in Figure 5.4, the relationship between reconstruction accuracy and hearing loss does seem to have become stronger after correction, even though the relationship at 500 Hz still seems to be reversed. These observations were reflected in the slopes of regression lines fit to the data at each frequency (Figure 5.5), with all slopes exhibiting a negative change (see Table 5.2). However, this relationship is still likely confounded by other factors, e.g., inter-frequency differences in reconstruction accuracy, i.e., the fact that the same degree of hearing loss at two different frequencies in the one subject can produce different

reconstruction accuracies, and it was decided that any further corrective action was unlikely to prove fruitful.

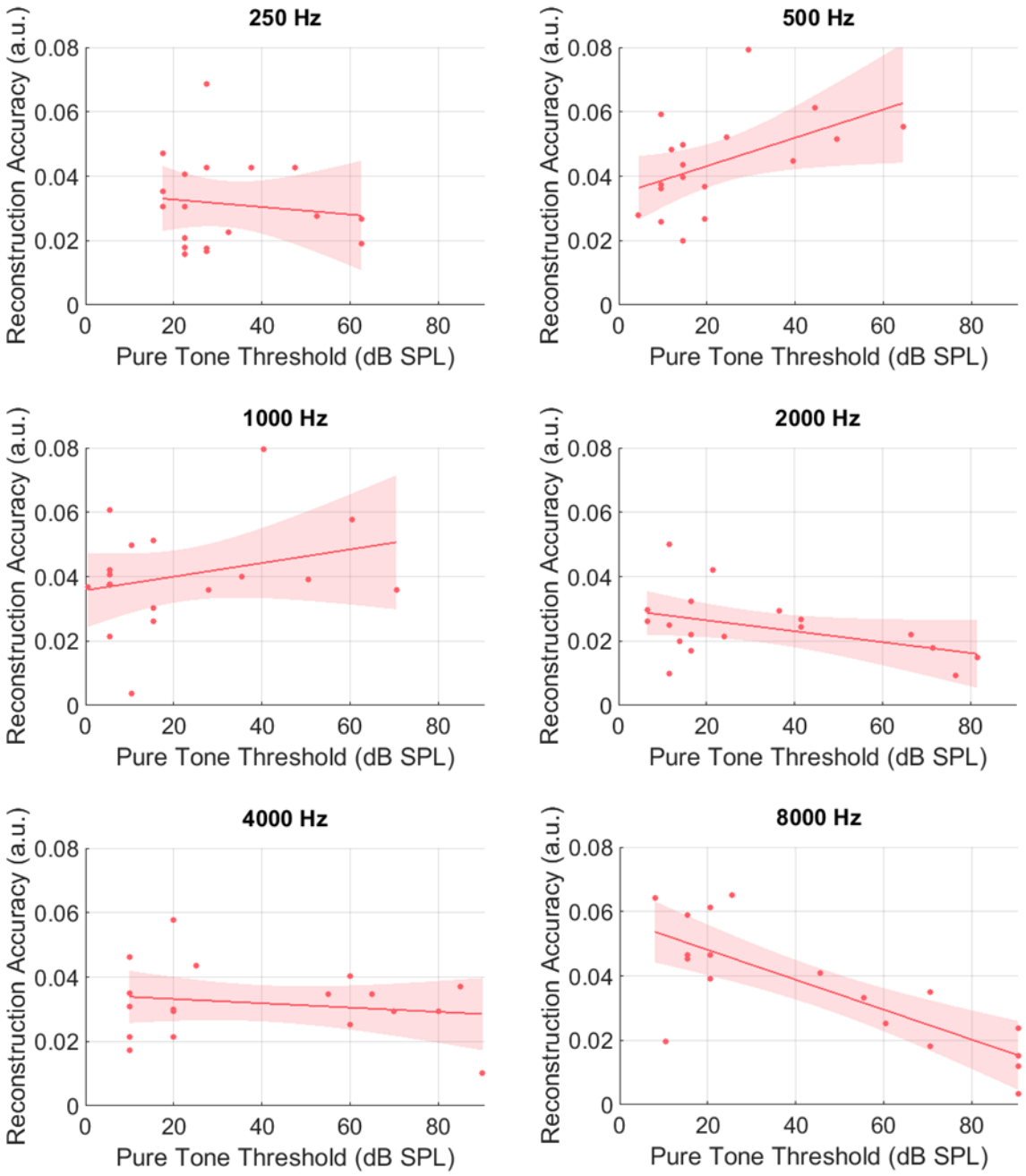

Figure 5.3: Scatter Plots of Uncompensated Reconstruction Accuracies against PT Thresholds at Each Frequency

Scatter plots of uncompensated reconstruction accuracies against PT thresholds at each frequency, with line fits and confidence intervals.

Table 5.1: Slopes of Regression Lines fit to Uncompensated Data at Each Frequency

Frequency (Hz)	Slope
250	-1.159×10^{-4}
500	4.394×10^{-4}
1000	2.115×10^{-4}
2000	-1.695×10^{-4}
4000	-6.684×10^{-5}
8000	-4.644×10^{-4}

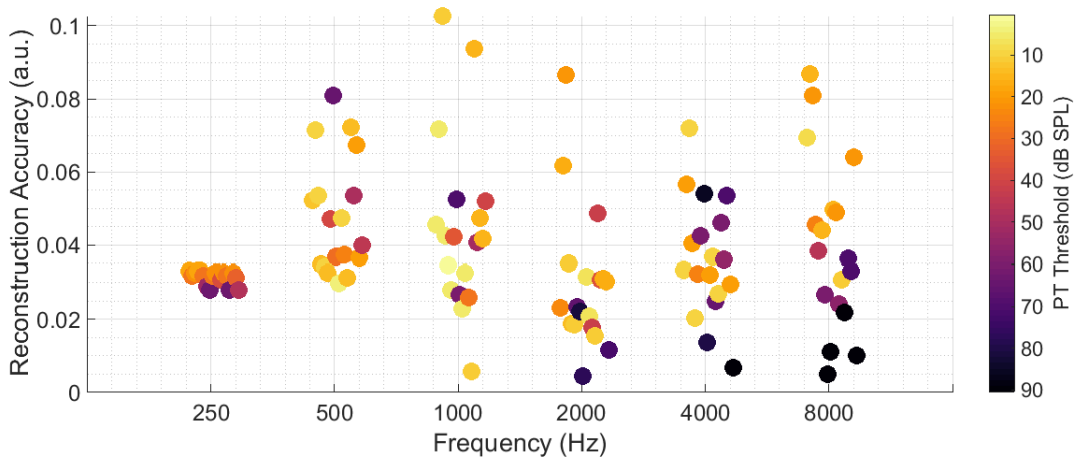

Figure 5.4: Compensated Global Relationship between Reconstruction Accuracy and Hearing Loss.

Reconstruction accuracies for each compensated AMT, for each subject, plotted against frequency and PT threshold. While the relationship between reconstruction accuracy and hearing loss does seem to have become stronger after correction, the relationship is still likely confounded by other factors, e.g., inter-frequency differences in hearing loss.

5.3.2 Using Reconstruction Accuracies to Determine Hearing Thresholds

Fortunately, the influence of individual and inter-frequency differences on reconstruction accuracy are limited when it comes to assessing an individual subject's, individual frequencies. However, these differences did mean that reconstruction accuracy alone was unlikely to be sufficient for determining hearing thresholds, and that another approach would be needed.

As mentioned, the overarching hypothesis of this work is that there is a relationship between reconstruction accuracy and hearing loss. Specifically, it was hypothesised that if a stimulus was heard by a normal hearing subject, one might expect almost all of the stimulus feature to have been encoded in the brain, resulting in a high reconstruction accuracy, but if the stimulus was heard by a hearing loss subject, one might expect much less of the stimulus feature to have been encoded in the brain, i.e., just the suprathreshold part, resulting in a lower reconstruction accuracy. One assumption inherent to this hypothesis—aside from the likely

lower SNR of the latter—is that how the stimulus feature is represented can influence the reconstruction accuracy.

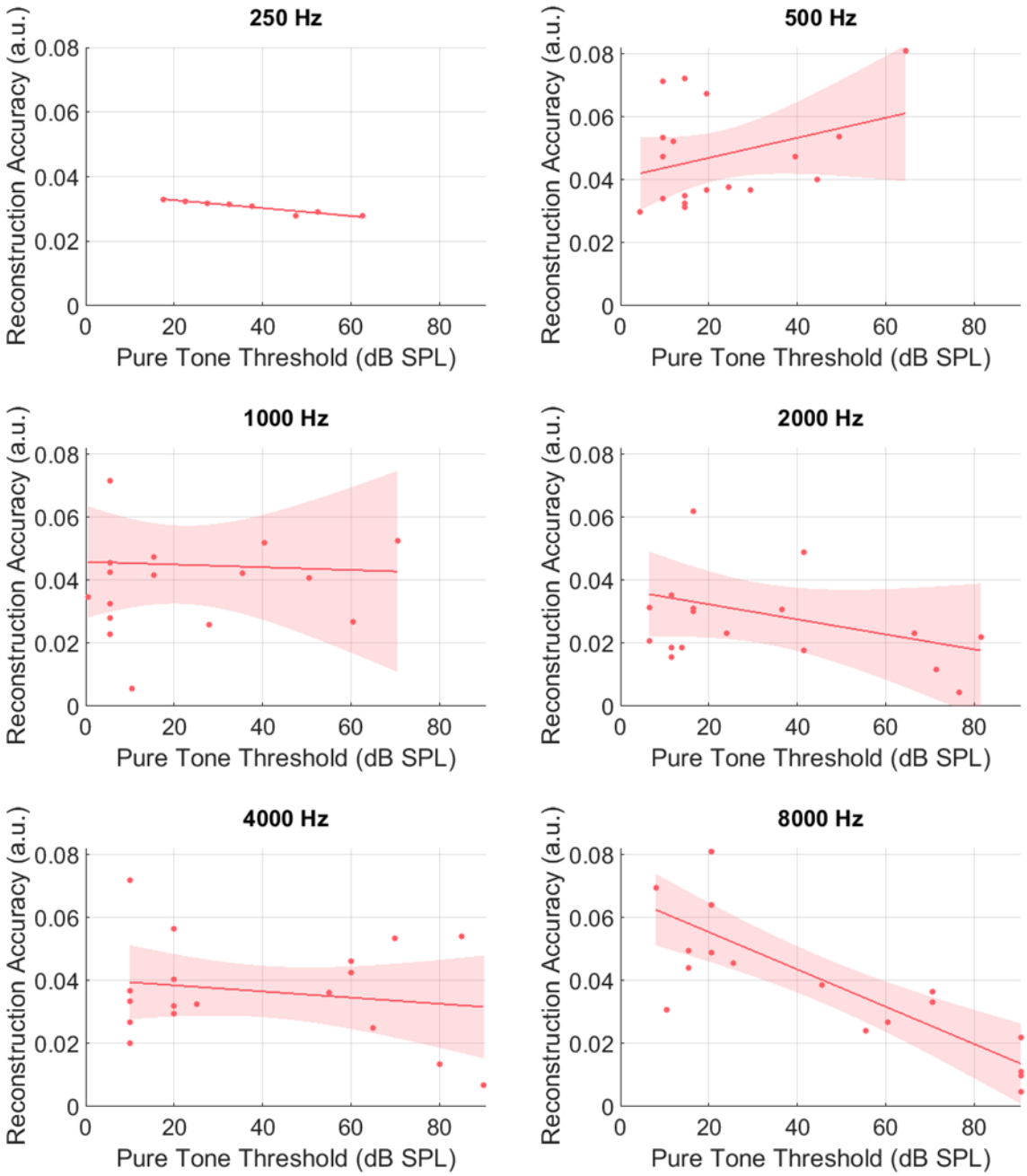

Figure 5.5: Scatter Plots of Compensated Reconstruction Accuracies against PT Thresholds at Each Frequency

Scatter plots of compensated reconstruction accuracies against PT thresholds at each frequency, with line fits and confidence intervals.

Table 5.2: Slopes of Regression Lines fit to Compensated Data at Each Frequency

Frequency (Hz)	Slope	Change
250	n/a	n/a
500	3.178×10^{-4}	-1.215×10^{-4}
1000	-4.360×10^{-5}	-2.551×10^{-4}
2000	-2.382×10^{-4}	-6.875×10^{-5}
4000	-9.786×10^{-5}	-3.102×10^{-5}
8000	-5.922×10^{-4}	-1.277×10^{-4}

For example, if a hearing loss subject only heard the upper (suprathreshold) part of the stimulus, and yet the full stimulus representation was used, one might expect a lower reconstruction accuracy than if a partial stimulus representation that better matched what they heard was used. This is because in the former case, even though only the suprathreshold part of the stimulus feature would likely be present in their EEG, the TRF would also try to fit the subthreshold part of the stimulus feature—as it would be present in the stimulus representation—likely adding noise to the model, resulting in a lower reconstruction accuracy. Corollary, if a normal hearing subject heard the full stimulus, and yet only the upper part of the stimulus representation was used, one might expect a lower reconstruction accuracy, than if a full stimulus representation that better matched what they heard was used. This is because in the former case, there would be variance in the EEG not accounted for by the model—as it would not be part of the stimulus representation—resulting in a lower reconstruction accuracy.

One way to test this hypothesis—and concurrently determine hearing thresholds—might be to iteratively derive models using stimulus representations that include ever increasing amounts of the stimulus feature, e.g., the top 5 dB, then the top 10 dB, and so on, until the full stimulus feature has been used. In theory then, as more and more relevant, i.e., heard, parts of the stimulus feature are included in the stimulus representation, more and more variance will be accounted for, increasing the reconstruction accuracy. Then, as the threshold is passed, and as more and more irrelevant, i.e., unheard, parts of the stimulus feature are included in the stimulus representation, more and more noise will be added to the model, decreasing the reconstruction accuracy. Therefore, if this hypothesis is true, the peak in reconstruction accuracy should occur at the subject's hearing threshold. This approach will heretofore be referred to as the moving-threshold approach, as it relies on simulating different hearing thresholds in the stimulus representation.

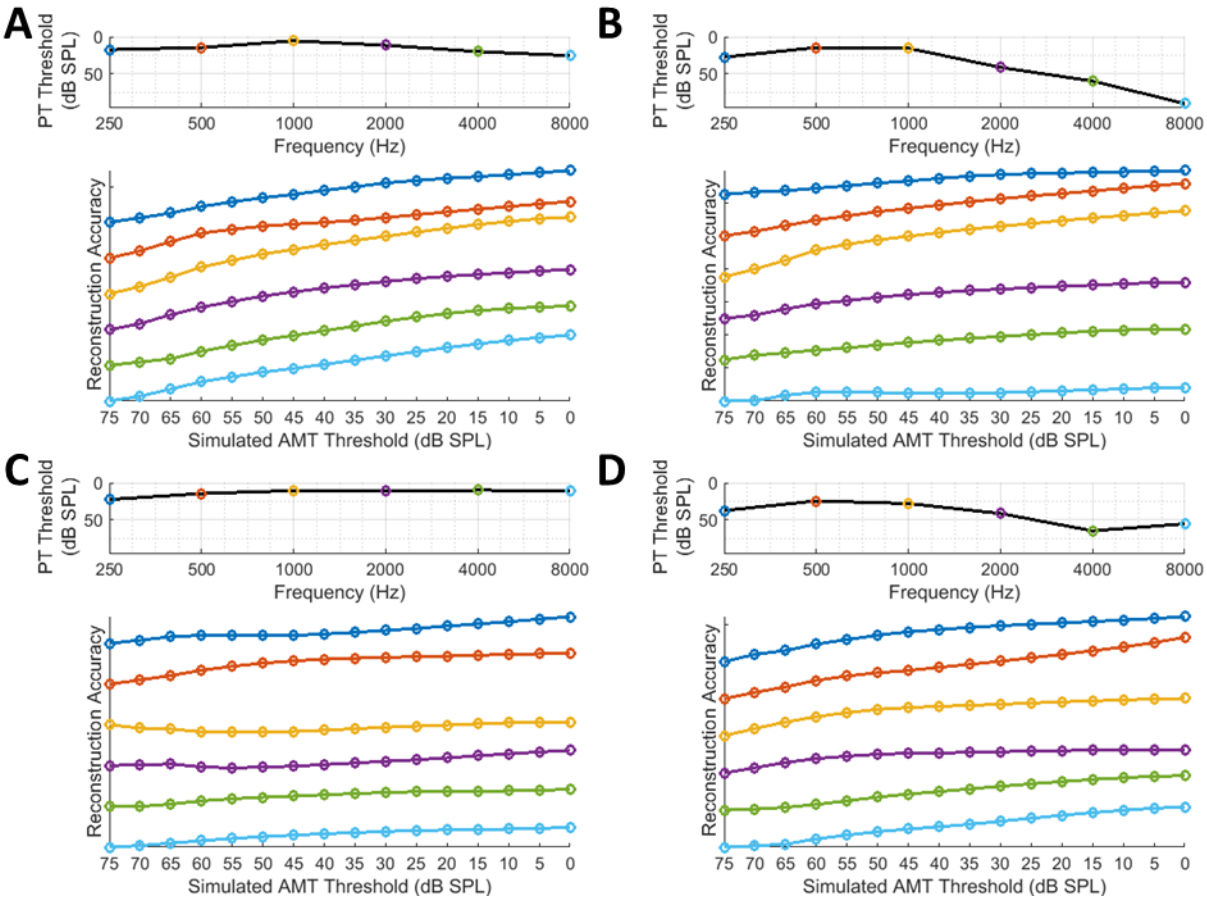

Figure 5.6: Representative Reconstruction Accuracy Profiles for the Moving-Threshold Approach

***A** – Reconstruction accuracy profiles for a normal hearing subject displaying rising trends. **B** – Reconstruction accuracy profiles for a hearing loss subject displaying rising trends but flatter trends for worse frequencies. **C** – Reconstruction accuracy profiles for a normal hearing subject displaying flatter and more variable trends. **D** – Reconstruction accuracy profiles for a hearing loss subject displaying rising and variable trends. Please note that the scale of the y-axis is inconsequential as the reconstruction accuracy profiles have been plotted stacked on top of one another to highlight their respective trends.*

Unfortunately, this hypothesis is not supported by the results. As can be seen from some representative data from two normal hearing subjects and two hearing loss subjects in Figure 5.6, the general trend was for reconstruction accuracies to rise continuously with stimulated AMT threshold, regardless of what their actual hearing threshold was. While it was expected that for normal hearing subjects, the reconstruction accuracies would rise across almost the entire range of simulated AMT thresholds—as they would not be expected to peak until almost the entire stimulus feature had been represented—the lack of a peak was not (e.g., Figure 5.6A). Furthermore, this rising trend was only seen in 6/9 of the normal hearing subjects, with the other 3 typically displaying flatter or more variable trends (e.g., Figure 5.6C). These rising

trends are perhaps even more conspicuous for the hearing loss subjects—as they would have been expected to peak much earlier—although flatter trends were seen at frequencies with higher thresholds for some subjects (e.g., Figure 5.6B). That said, this was only seen in 3/9 of the hearing loss subjects, with the other 6 typically displaying rising or more variable trends (e.g., Figure 5.6D).

5.4 Discussion

We introduced a novel approach for objectively determining hearing thresholds using TRF estimation and MAMTs. While we were unable to actually determine hearing thresholds using this approach, our analyses did seem to indicate that there is a relationship between reconstruction accuracy and hearing loss. In this section, we will discuss the results of our analyses, and outline a number of methods that could be used to improve the current approach.

Our initial analyses indicated, that there did seem to be a relationship between reconstruction accuracy and hearing loss (Figure 5.2 and Figure 5.3). It was clear from these early efforts, however, that this relationship was likely confounded by individual differences in reconstruction accuracy. These could be due to individual differences in envelope tracking, i.e., how well the AMT envelope is tracked in the brain, neural coding, i.e., how much of the envelope tracking is temporally coded vs. rate coded, cortical folding, i.e., how the subject's cortex is folded, which can affect the projection of electric fields towards the scalp, and thus likely SNR, etc. While subsequent efforts to account for some of these differences did seem to make this relationship stronger (Figure 5.4 and Figure 5.5), it also seemed likely that this relationship was still confounded by other factors, e.g., inter-frequency differences in reconstruction accuracy not simply due to differences in threshold. These could be caused by similar processes as previously, except that they differ on a frequency-to-frequency basis. It was, therefore, decided that the potential benefits of this line of enquiry were limited and that another approach should be taken.

Fortunately, the influence of individual and inter-frequency differences in reconstruction accuracy are limited when it comes to assessing an individual subject's, individual frequencies, and the introduction of the moving-threshold approach provided a more level-sensitive measure with which to determine hearing thresholds. To reiterate, the hypothesis behind the moving-threshold approach, was that as more and more relevant parts of the stimulus feature were included in the stimulus representation, the reconstruction accuracy

would increase, and as more and more irrelevant parts of the stimulus feature were included in the stimulus representation, the reconstruction accuracy would decrease, with the peak in reconstruction accuracy occurring at the subject's hearing threshold. Unfortunately, this hypothesis was not supported by the results, with most reconstruction accuracies displaying a rising, variable, or flat trend (e.g., Figure 5.6). It is not clear why this would have been the case, particularly for the hearing loss subjects for whom the peak should have occurred much earlier. In future work, it may be beneficial to first simulate hearing loss in normal hearing subjects—by presenting more or less of the stimulus—to try and get a better understanding of the relationship between reconstruction accuracy and (simulated) hearing loss, as well as the influence of stimulus representation in a more controlled experimental environment.

Another approach that might prove useful could be to use forward mapping and response detection, like many other objective approaches (Van Maanen and Stapells, 2005). While forward mapping is sub-optimal—in that it does not exploit all the available data when deriving stimulus-response mappings—it does come with the benefit of neurophysiological interpretability. Indeed, one of the shortcomings of the ASSR approach outlined earlier, was the lack of temporal resolution in the response. This could also potentially make many more measures available for assessment, such as peak-to-peak amplitudes, latencies, etc. That said, it is not clear that the moving threshold approach would be the optimal way to exploit these data, and the development of other approaches may be required.

Efforts could also be made to move towards a more level-adaptive approach, as used by many other objective approaches (Elberling et al., 2007a). This could be achieved by changing the amplitude distribution of the stimulus as soon as a response has been detected at a particular level. This would mean that more time would be spent stimulating at the levels most relevant to threshold determination, hypothetically improving the SNR in that region and reducing the testing time. That said, this type of approach would not be well suited for cases where both low- and high-level processing were being assessed simultaneously, e.g., if using speech, but should suffice if threshold determination is the main concern. Another consideration if using speech, is that the frequency-specific envelopes would be highly correlated, and so it would be important that all the envelopes be represented in a multivariate way, when fitting the models. This would be challenging with the current moving-threshold approach, given the large number of combinations that would need to be tested, and so again, the development of other approaches to exploit this type of data may be required.

In summary, we have attempted to introduce a novel approach for objectively determining hearing thresholds using TRF estimation and MAMTs. While we were unable to actually determine hearing thresholds using this approach, our analyses did seem to indicate that there is a relationship between reconstruction accuracy and hearing loss, i.e., in general, the worse the hearing loss the lower the reconstruction accuracy. Future work could focus on studying this relationship in a more controlled experimental setting and evaluating the potential of the proposed forward modelling and response detection approach.

Chapter 6. Stimulus Dependent Modelling of the Cortical Tracking of Complex Sound Envelopes

6.1 Introduction

In this chapter, several novel stimulus representations are presented with the view of enhancing response derivation using TRF estimation. The importance and benefits of taking certain neurophysiological properties of the human auditory system into account when designing stimulus representations are discussed and then quantified through comparisons with other models derived using more standard stimulus representations. A manuscript on this study has been published in *eNeuro* (Drennan and Lalor, 2019).

Implicitly, TRFs assume that that the auditory system is linear, and that responses to a stimulus feature can be modelled by a linear impulse response function. In other words, they assume that responses to a stimulus feature will be temporally and morphologically consistent across its time-course, but just scaled linearly as a function of the stimulus feature's intensity. While this assumption may be reasonable for certain brain responses, in certain brain areas, to certain stimulus features (Boynton et al., 1996), there is definitive evidence that it is imperfect for EEG-based TRFs.

One such piece of evidence is the long-known relationship between auditory stimulus amplitude and response latency (Beagley and Knight, 1967). Specifically, while there is a monotonic—although not necessarily linear—relationship between auditory stimulus amplitude and response magnitude, there is also an inverse relationship between stimulus amplitude and response latency. Therefore, to model neural responses to an ongoing auditory

stimulus using a linear univariate TRF is likely to be suboptimal given that it ignores the dependence of response latency—and morphology—on stimulus amplitude.

Here, we aim to demonstrate that by allowing the stimulus-response model to vary as a function of the stimulus amplitude, we can improve the modelling of responses to continuous auditory stimuli. To do so, we propose a simple extension to the standard linear TRF estimation approach that involves amplitude binning a *single* feature, namely the envelope, and then using it to fit a *multivariate* TRF. This should allow the TRF to vary across the different amplitude ranges, thus enabling it to account for associated changes in response magnitude, latency, and morphology. We aim to validate that this represents an improved model by comparing how well it predicts EEG data relative to more standard univariate models.

Furthermore, we aim to demonstrate that this model can be further improved by including an additional envelope representation that emphasises onsets and positive changes in the stimulus, in line with the idea that while some neurons track the entire envelope, others respond preferentially to onset, offsets, and changes in the stimulus (Bieser and Müller-Preuss, 1996). These improved models could provide a more accurate and detailed view of the auditory system under study, potentially leading to a deeper understanding of how the auditory system responds to continuous stimuli, and more sophisticated measures of low-level auditory processing.

6.2 Materials and Methods

EEG data from two experiments were used in this study: one acquired in response to AM BBN, i.e., the data collected for Experiment 2 in Chapter 4, the other in response to continuous natural speech (Natural Speech Dataset from https://doi.org/10.5061/dryad.070jc including amplitude envelopes; Broderick et al., 2018).

6.2.1 Subjects

13 subjects aged 23–35 years participated in the AM BBN experiment; 5 were male. 19 subjects aged 19–38 years participated in the speech experiment; 13 were male, although data from 2 subjects were later excluded because of uncertainties in response timing due to differences in their data acquisition setup. All subjects had self-reported normal hearing. The protocols for both studies were approved by the Ethics Committee of the Health Sciences Faculty at Trinity College Dublin, Ireland, and all subjects gave written informed consent.

6.2.2 Stimuli

As mentioned, this study involved experiments using two different types of stimuli, AM BBN and continuous natural speech.

The carrier signal for the AM BBN stimulus was uniform BBN with energy limited to a bandwidth of 0–24000 Hz. Its modulating signal (envelope) had a log-uniform amplitude distribution—by design, although less so after envelope extraction (see below)—and a bottom-heavy (right-skewed) frequency (modulation rate) distribution (Figure 6.1). The envelope was created by first generating a signal with discrete amplitude values with the desired statistical properties, and then interpolating between those discrete points to provide a smooth transition from one modulation amplitude to the next.

The speech stimulus had a bottom-heavy (right-skewed) frequency distribution with energy limited to a bandwidth of 0–22050 Hz. Its envelope had a log-top-heavy (left-skewed) amplitude distribution, and a bottom-heavy (right-skewed) frequency (modulation rate) distribution, similar to that of the AM BBN stimulus (Figure 6.1). It comprised extracts from a professional audio-book version of a popular mid-20th century American work of fiction, i.e., The Old Man and the Sea by Ernest Hemingway, written in an economical and understated style and read by a single male American speaker.

6.2.3 Experimental Procedure

In the AM BBN experiment, subjects were presented with 80 repetitions of the same 60 s long AM BBN stimulus as they reclined in a comfortable chair, in a quiet, darkened room, and watched a silent animated cartoon presented on a tablet computer. They were asked *not* to attend to the auditory stimuli, which were presented monaurally to their right ear at a peak level equivalent to that of a 1 kHz PT at 80 dB SPL, using a Sound Blaster X-Fi Surround 5.1 Pro external sound card, a TPA3118D2EVM amplifier, and electromagnetically shielded Etymotic Research ER-2 earphones, via VLC Media Player from VideoLan (http://www.videolan.org). Compensation for the 1 ms sound tube delay introduced by the ER-2 earphones was applied post-hoc.

In the speech experiment, subjects were presented with 28 trials of ~155s long audiobook extracts. The trials preserved the storyline, with neither repetitions nor discontinuities. Subjects sat in a comfortable chair, in a quiet, darkened room, and were instructed to maintain visual fixation on a crosshair centred on a computer monitor, and to

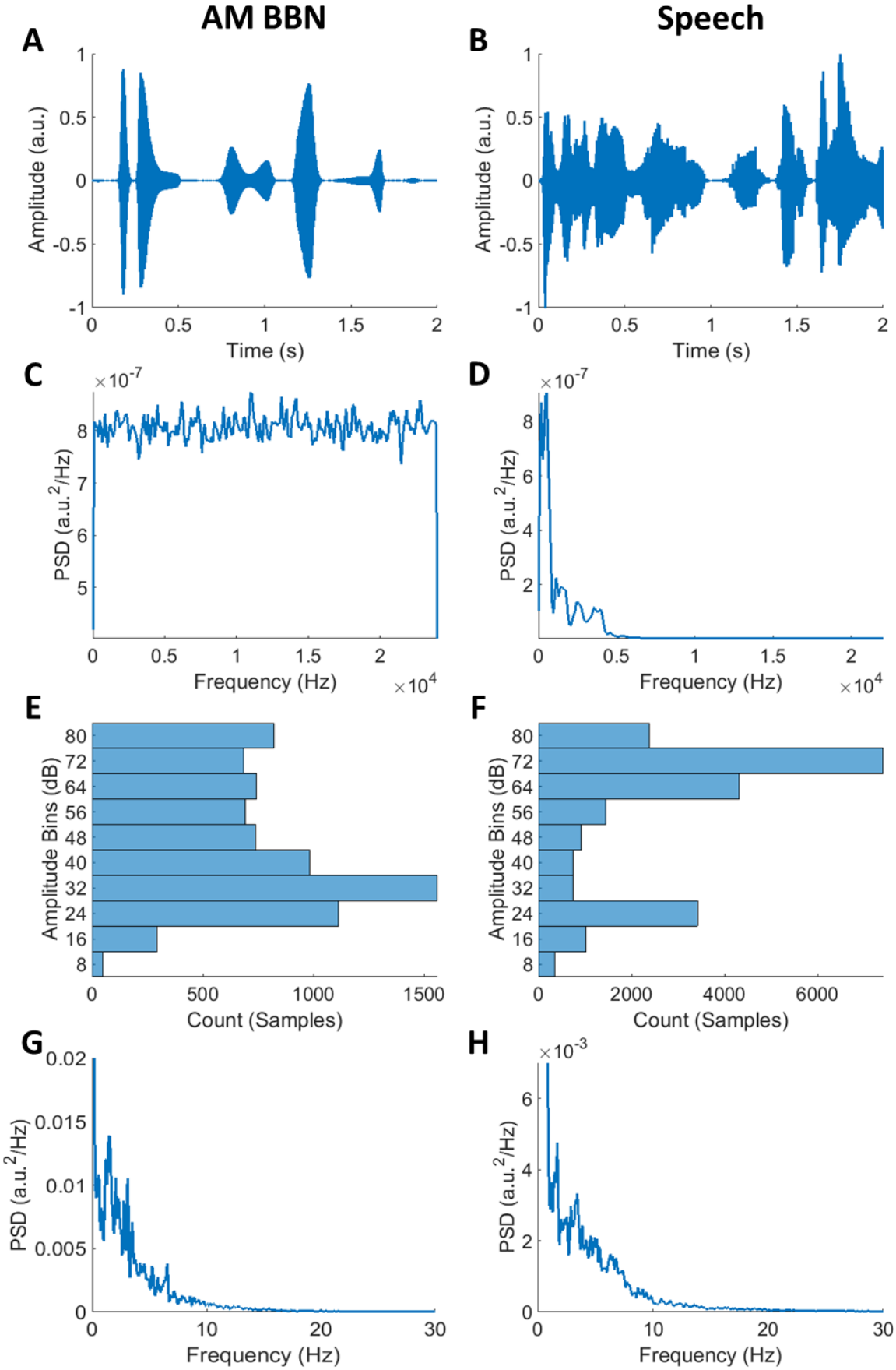
AM BBN
Speech
A
B
Amplitude (a.u.)
Time (s)
C
D
PSD (a.u.^{2}/Hz)
Frequency (Hz)
$\times 10^{-7}$
$\times 10^{4}$
E
F
Amplitude Bins (dB)
Count (Samples)
G
H
$\times 10^{-3}$

Figure 6.1: Example Segments and Properties of the Stimuli Used in This Study

A,B *– Example segments of AM BBN and speech stimuli, respectively.* ***C,D*** *– Power spectral densities (PSDs) of AM BBN and speech stimuli, respectively. The AM BBN had a broadband frequency distribution by design, while the male speaker had a frequency distribution that was dominated by frequencies below 5000 Hz.* ***E,F*** *– Amplitude histograms of AM BBN and speech envelopes, respectively. Both envelopes had quite broadly distributed amplitude distributions. Please note that the amplitude distribution of the AM BBN envelope was uniform by design, but after extracting the envelope from the AM BBN signal using the Hilbert transform, it was less so. Please also note that the amplitude distribution of the speech envelope was more skewed, with a higher percentage of samples in the higher amplitude bins.* ***G,H*** *– PSDs of AM BBN and speech envelopes, respectively. Both signals had envelopes with a bottom-heavy (right-skewed) frequency distribution indicating that their modulation rates were dominated by low frequencies.*

minimise eye blinking and all other motor activities for the duration of each trial. They were asked to attend to the auditory stimuli, which were presented diotically at a comfortable listening level, using Sennheiser HD 650 headphones, via Presentation software from Neurobehavioral Systems (http://www.neurobs.com). For the purposes of analysis, all trials were truncated to 150 s, and a peak level of 80 dB SPL was estimated—as the original presentation level was not available.

6.2.4 EEG Acquisition

In the AM BBN experiment, 40 channels of EEG data were recorded at 16384 Hz (analog -3 dB point of 3276.8 Hz), using a BioSemi ActiveTwo system (http://www.biosemi.com). 32 cephalic electrodes were positioned according to the standard 10-20 system. A further eight non-cephalic electrodes were also collected although only two—those over the left and right mastoids—were used in the analysis. Triggers indicating the start of each 60 s trial were encoded in a separate channel in the stimulus WAV file as three cycles of a 16 kHz tone burst. These triggers were interpreted by custom hardware before being fed into the acquisition laptop for synchronous recording along with the EEG.

In the speech experiment, 130 channels of EEG data were recorded at 512 Hz (analog -3 dB point of 409.6 Hz), using a BioSemi ActiveTwo system. 128 cephalic electrodes were positioned according to the BioSemi Equiradial system, with another 2 electrodes located over the left and right mastoids. Triggers indicating the start of each ~155 s trial were presented using Neurobehavioral Systems Presentation software for synchronous recording along with the EEG.

6.2.5 EEG Preprocessing

The EEG data were first resampled to 128 Hz using the *decimate* function in MATLAB (http://www.mathworks.com) and implemented using the *filtfilt* function. A 1st order high-pass Butterworth filter was then applied with a cutoff frequency of 1 Hz, also using the *filtfilt* function. Bad channels were determined as those whose variance was either less than half or greater than twice that of the surrounding 2–4 channels for the AM BBN dataset, and 3–7 channels for the speech dataset, depending on location. These were then replaced through spherical spline interpolation using EEGLAB (Delorme and Makeig, 2004). Finally, the data were rereferenced to the average of the mastoids, separated into trials based on the triggers provided, and z-scored.

6.2.6 Temporal Response Function Estimation

Responses were derived using TRF estimation, and were implemented via the mTRF Toolbox (Crosse et al., 2016) using a forward modelling approach (see Chapter 3 for details). Baseline correction was performed on each subject's average TRF—by subtracting the mean value between -20 and 0 ms—before being combined to form the grand average response.

6.2.7 Amplitude-Binned Envelope Stimulus Representation

As has been mentioned throughout this thesis, the choice of stimulus feature can have a significant influence on the resulting model. Such features could include the envelope (Lalor et al., 2009) or spectrogram (Ding and Simon, 2012; Di Liberto et al., 2015), or in the case of speech, phonemes, phonetic features (Di Liberto et al., 2015), or its semantic content (Broderick et al., 2018). The envelope (time x amplitude), however, is probably the most commonly used stimulus feature and is the one chosen for use in this study. For both the AM BBN and speech stimuli the envelopes were calculated by taking the absolute value of their Hilbert transforms, and then resampling them to 128 Hz using the *decimate* function in MATLAB.

As mentioned, it has long been known that the magnitude and latency of auditory system responses vary directly and inversely with stimulus amplitude, respectively, i.e., as the stimulus amplitude increases, the response magnitude increases, and the response latency decreases—and vice versa. Univariate TRFs, like those modelled using envelopes, cannot account for all these amplitude-dependent changes. In fact, univariate TRFs can only account

for linear changes in magnitude and cannot account for any changes in latency or morphology. However, by simply amplitude binning the envelope (time x [amplitude] x amplitude), i.e., by dividing the envelope up into multiple sub-envelopes comprising the different amplitude ranges of the full envelope, normalising the values in each bin to be between 0 and 1, and then using it to fit a multivariate TRF, should allow the TRF to vary across the different amplitude ranges, potentially enabling it to account for more of these amplitude-dependent changes than its univariate counterpart.

The amplitude binned (AB) envelope was created by logarithmically binning the envelope into 8 dB bins using the *histcounts* function in MATLAB, and then normalising the values in each bin to between 0 and 1—an important step in ensuring the stability of the resulting TRF. This bin size was chosen empirically after comparing the prediction accuracies attained across a range of bin sizes, with broader bins perhaps being less able to capture changes in the response with amplitude, and narrower bins perhaps suffering from the limited amount of data available for training. The logarithmic bin edges were determined by taking 10 to the power of the desired bin edges in dB, i.e., 8, 16, 24, etc., divided by 20, and then normalising the resulting range to between 0 and 1 (Figure 6.2B).

6.2.8 Other Stimulus Representations

A number of other approaches have already been put forward that attempt to modify the stimulus representation in order to account for certain properties of the auditory system. So, rather than just comparing the AB envelope model with the standard envelope model, we also chose to compare it with two others, i.e., the SPL envelope and onset envelope models. The SPL envelope model was fit using an envelope that was transformed into its equivalent logarithmic SPL representation, and the onset envelope model was fit using an envelope that was modified to place a greater emphasis on onsets and positive changes in amplitude.

The motivation for using the SPL envelope model derives from the well-known fact that electrophysiological responses generally vary in proportion to the log of the stimulus amplitude (Aiken and Picton, 2008). The SPL envelope was generated by taking 20 times the base 10 logarithm of the envelope (Aiken and Picton, 2008; Figure 6.2A), and it was hoped that this would help linearise the amplitude to magnitude mapping between the stimulus representation and the EEG. It was presumed that the AB envelope model might outperform the SPL envelope model, however, given that they both attempt to account for nonlinearities

in the relationship between stimulus amplitude and response magnitude, but only the former accounts for changes in response latency and morphology.

The motivation for using the onset envelope model comes from the idea that many auditory neurons are particularly sensitive to onsets, offsets, and changes in the stimulus (Bieser and Müller-Preuss, 1996), and that this approach has been used effectively in the past (Aiken and Picton, 2008; Hertrich et al., 2012; Fiedler et al., 2017). The onset envelope was explicitly designed to reflect onsets and positive changes in the stimulus, and was created by half-wave rectifying the first-derivative of the envelope (Hertrich et al., 2012; Figure 6.2A). This is equivalent to applying a high-pass filter with a cutoff frequency of $\frac{Fs}{4}$, i.e., 32 Hz.

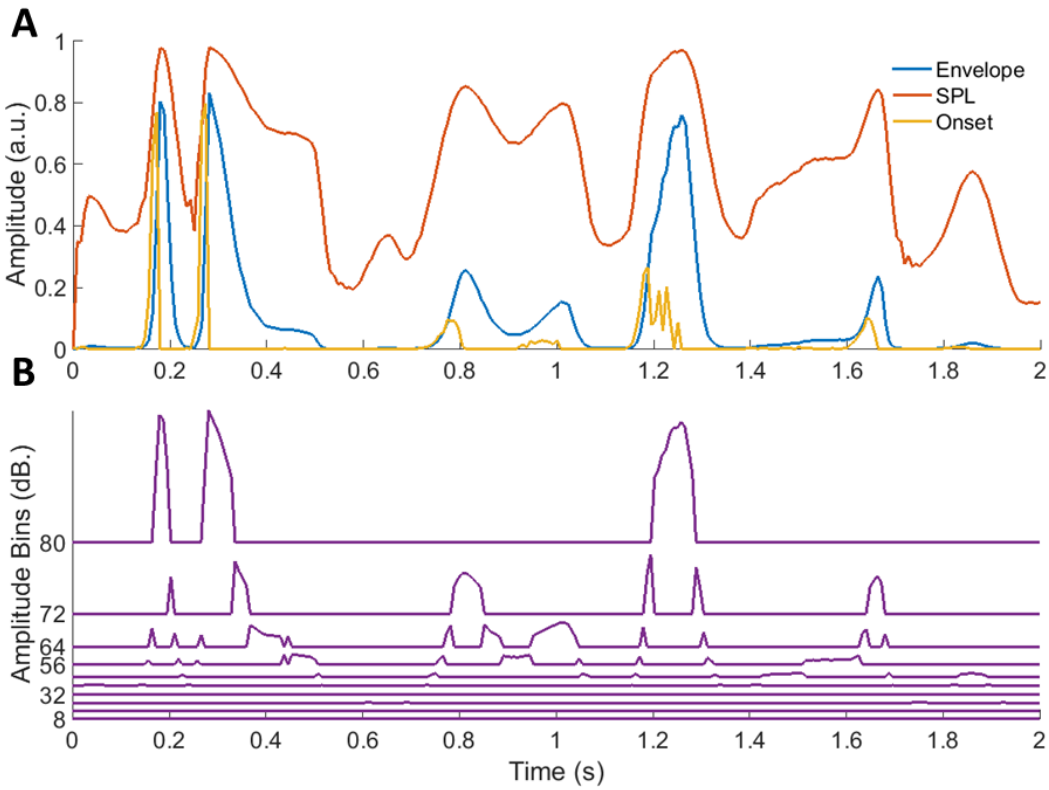

Figure 6.2: Example Segments of Some of the Stimulus Representations Used in this Study

A *– Example segments of the envelope, SPL envelope, and onset envelope stimulus representations.* ***B*** *– Corresponding segment of the AB envelope.*

6.2.9 Model Comparison

In order to compare the different models tested as part of this study, a nested leave-one-out cross-validation approach was employed. Specifically, for each stimulus representation, a separate TRF—univariate for the envelope, SPL envelope, and onset envelope, and multivariate for the AB envelope—was fit for each of M trials across several ridge parameters

used to regularise the models. One trial was then chosen to be left out, i.e., to be used as a test set, with the remaining M-1 trials to be used for the inner cross-validation. Of these inner M-1 trials, one trial was again chosen to be left out, i.e., to be used as a validation set, with the remaining M-2 trials to be used as a training set.

For each lambda value, an average model was obtained by averaging over the single-trial models in the training set. These were then convolved with the stimulus representation associated with the validation set to predict its neural response. Model performance was assessed by quantifying how accurately these predicted responses matched the actual recorded response from the validation set, using Pearson's correlation coefficient. This process was then repeated M-2 times such that each trial was left out of the training set once. The overall model performance was then determined by averaging over the individual model performances for each trial, and the optimal lambda value was chosen.

Using this optimal lambda value, another average model was then obtained by averaging over the single-trial models in both the training and validation sets. This was then convolved with the stimulus representation associated with the test set to predict its neural response. Model performance was then assessed by quantifying how accurately the predicted response matched the actual recorded response from the test set. This entire procedure was then repeated M-1 times such that each trial was left out of the inner cross-validation procedure once. The overall model performance was then finally determined by averaging over the individual model performances for each trial. Importantly, the parameter optimisation was done separately for each stimulus representation and subject, so that we were left comparing each model based on its respective optimal performance.

Again, the performance of each model was assessed by quantifying how accurately the predicted response matched the actual recorded response, using Pearson's correlation coefficient. The normality of these performance measures for each model was confirmed using the Anderson Darling test, and model comparisons were carried out using paired sample t-tests and Cohen's d effect size for paired sample t-tests. Cohen's d effect size was calculated by dividing each t-value by the square root of the sample size. One potential concern when comparing models with different numbers of parameters is that models with more parameters may perform better simply due to their greater complexity. To account for this, supplementary comparisons were also carried out using the Akaike Information Criterion (AIC) which penalises models based on their complexity. As the results of these analyses were not normal, model comparisons were carried out using Wilcoxon signed-rank tests.

Permutation tests were also used to assess the null distributions of the envelope models. For the AM BBN dataset, as the stimulus was the same for each trial, a pool of 80 circularly-shifted envelopes, i.e., the original envelope plus 79 circularly-shifted envelopes, each iteratively shifted by 1/80 times the length of the envelope with respect to the previously shifted envelope, were first created. 80 envelopes from this pool were then chosen at random with replacement for use in the cross-validation procedure. This selection and cross-validation procedure was repeated 100 times to determine the null-distribution of the envelope model for each subject. For the speech dataset, as the stimuli were different for each trial, envelopes were simply chosen at random with replacement from the original set of envelopes, for use in the cross-validation procedure. This selection and cross-validation procedure was also repeated 100 times to determine the null-distribution of the envelope model for each subject.

6.3 Results

6.3.1 Channel Selection

EEG prediction accuracies will vary across channels depending on how related the data on those channels are to the stimulus representation. For the AM BBN analyses, the seven channels (of 32) with the highest prediction accuracies for the envelope model were used (Figure 6.3A). For the speech analyses, the 42 channels (of 128) with the highest prediction accuracies for the envelope model, plus three other channels —to ensure symmetry—were used (Figure 6.5A). In both cases, these channels tended to reside over fronto-central to temporal scalp (see Di Liberto et al., 2015). The overall prediction accuracy for each model, i.e., the accuracy with which unseen EEG can be predicted by each model, was calculated by averaging the prediction accuracies over these electrodes.

6.3.2 Individual Model Comparisons

<u>AM BBN</u>

For the AM BBN dataset, prediction accuracies were determined for each model, and each subject (Figure 6.3B). All four stimulus representations, i.e., envelope, SPL envelope, onset envelope, and AB envelope, and their associated models were able to predict EEG responses with an accuracy that was significantly above 0.0012, i.e., the null hypothesis obtained using the permutation tests, for all subjects ($t(12)$, all $p < 0.001$), and greater than all values obtained

using the permutation tests. However, the AB envelope model significantly outperformed all three of the other models, in each case with a large to very large positive effect size (see Table 6.1). These results were also seen when comparing the models using AIC ($p < 0.001$; Wilcoxon signed rank test). Neither the SPL envelope nor onset envelope models managed to outperform the standard envelope model ($t(12)$, both $p > 0.05$).

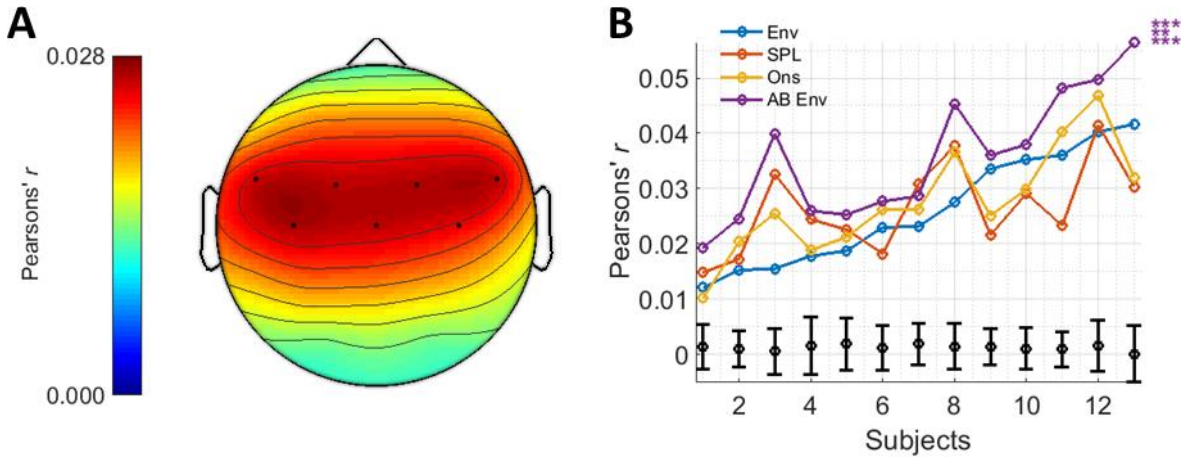

Figure 6.3: Prediction Accuracies for the AM BBN Dataset

A *– Topographic plot displaying prediction accuracies for the envelope model for the AM BBN dataset and highlighting the channels chosen for analysis.* ***B*** *– Prediction accuracies for each model and subject, including null hypotheses for the envelope model as determined from the permutation tests, and indications of significance as determined from the t-tests.*

Table 6.1: Comparison Between AB Envelope and Other Models for the AM BBN Dataset

Model	df	t	p	d
Envelope	12	5.471	< 0.001	1.518
SPL Envelope	12	4.070	< 0.01	1.129
Onset Envelope	12	4.800	< 0.001	1.331

Exactly how the TRF changes as a function of stimulus amplitude becomes more apparent on closer inspection of the AB envelope TRF (Figure 6.4A). As the stimulus amplitude decreases, the TRF magnitude decreases, latency increases, and morphology changes in accordance with our hypothesis. The influence of stimulus amplitude on TRF latency is perhaps better emphasised in Figure 6.4B. For example, the "N1", which is quite large in magnitude in the uppermost amplitude bin, decreases in magnitude and increases in latency, with decreasing stimulus amplitude. To quantify this relationship, the N1 peak in each bin was determined as being the largest negative peak in the TRF at lags between 70 and 210 ms (the corresponding latencies can be seen in Figure 6.4D). A line was then fit to the data (R-squared = 0.9143, $p < 0.001$), which showed that the N1 peak latency increases by ~11 ms with

every unit decrease in amplitude bin. A similar effort was made to quantify the relationship between stimulus amplitude and TRF magnitude, i.e., the "P1" peak in each bin was determined as being the largest positive peak in the TRF at lags between 0 and 130 ms, and the corresponding P1-N1 peak-peak amplitudes can be seen in Figure 6.4F. However, while there does seem to be some relationship between stimulus amplitude and TRF magnitude, it was not well fit by a line (R-squared = 0.597, $p < 0.01$).

Speech

For the speech dataset, prediction accuracies were again determined for each model, and each subject, with very similar results to before (Figure 6.5B). All four stimulus representations and their associated models were able to predict EEG responses with an accuracy that was significantly above 0.0015, i.e., the null-hypothesis obtained using the permutation tests, for all subjects (t(16), all $p < 0.001$), and greater than all values obtained using the permutation tests. The AB envelope model significantly outperformed all three of the other models, in each case with a large to very large positive effect size (see Table 6.2). These results were also seen when comparing the models using AIC ($p < 0.001$; Wilcoxon signed rank test). Neither the SPL envelope nor onset envelope models managed to outperform the standard envelope model (t(16), both $p > 0.05$).

Again, exactly how the TRF changes as a function of stimulus amplitude becomes more apparent on closer inspection of the AB envelope TRF (Figure 6.6A and B). While the overall relationship between stimulus amplitude and TRF magnitude, latency, and morphology appears similar to before, in this case, the magnitude of the TRF for some of the lower amplitude bins seems unexpectedly high. It is not entirely clear why this would have been the case. The "P1" and "N1" peaks were also determined in the same manner as before, and the corresponding N1 latencies and P1-N1 peak-peak amplitudes can be seen in Figure 6.6D and F respectively. To quantify the relationship between stimulus amplitude and TRF latency, a line was fit to the N1 latency data (R-squared = 0.5008, $p < 0.05$), which again showed that the N1 peak latency increases by ~11 ms with every unit decrease in amplitude bin. However, there was no simple relationship between stimulus amplitude and TRF magnitude.

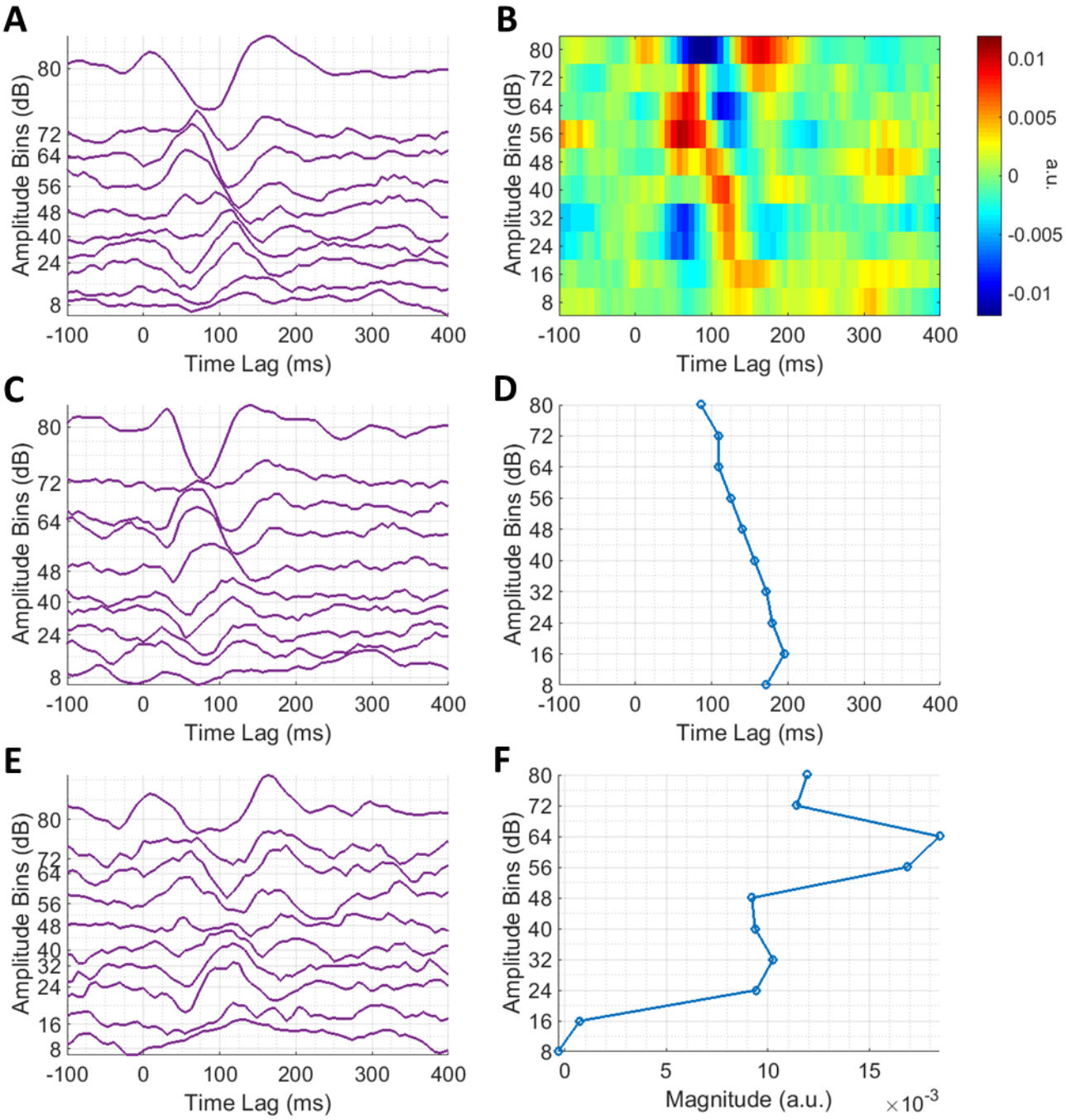

Figure 6.4: Analysis of Amplitude-Dependent Changes for the AM BBN Dataset.

A *– Group average AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***B*** *– Image plot of group average AB envelope TRF.* ***C*** *– Single subject AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***D*** *– N1 peak latencies across group average AB envelope TRF bins.* ***E*** *– Single subject AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***F*** *– P1-N1 peak-to-peak amplitudes across group average AB envelope TRF bins. Please note, these findings represent amplitude-dependent changes at a single representative channel over left central scalp for the AM BBN stimulus.*

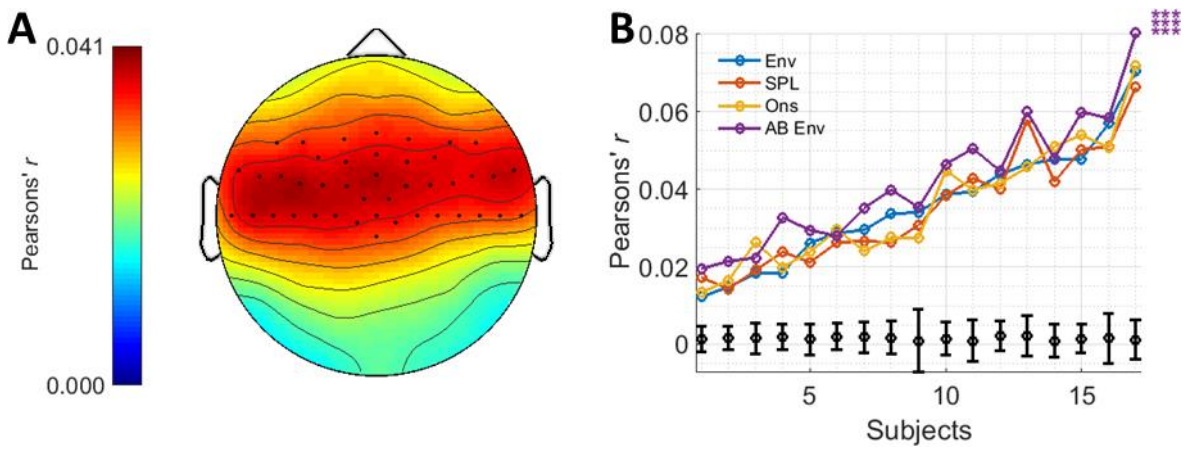

Figure 6.5: Prediction Accuracies for the Speech Dataset

A *– Topographic plot displaying prediction accuracies for the envelope model for the speech dataset and highlighting the channels chosen for analysis.* ***B*** *– Prediction accuracies for each model and subject, including null hypotheses for the envelope model as determined from the permutation tests, and indications of significance as determined from the t-tests.*

Table 6.2: Comparison Between AB Envelope and Other Models for the Speech Dataset

Model	df	t	p	d
Envelope	16	5.472	< 0.001	1.327
SPL Envelope	16	7.649	< 0.001	1.855
Onset Envelope	16	4.666	< 0.001	1.132

6.3.3 Combined Model Comparisons

The comparison between the AB envelope and onset envelope models is not necessarily as straightforward as one might expect, however. This is because each model is likely reflecting different envelope tracking mechanisms in the cortex (Bieser and Müller-Preuss, 1996). Specifically, the onset envelope model likely reflects contributions from neurons that track onsets and positive changes in amplitude while the AB envelope model likely reflects contributions from neurons that track along with all of the amplitude fluctuations (Bieser and Müller-Preuss, 1996).

To test the idea that these two models are capturing complementary information on envelope tracking, we investigated whether there would be any advantage in combining these two models, i.e., by combining the two stimulus representations and then using that to fit a multivariate TRF. Indeed, the combined AB envelope plus onset envelope model significantly

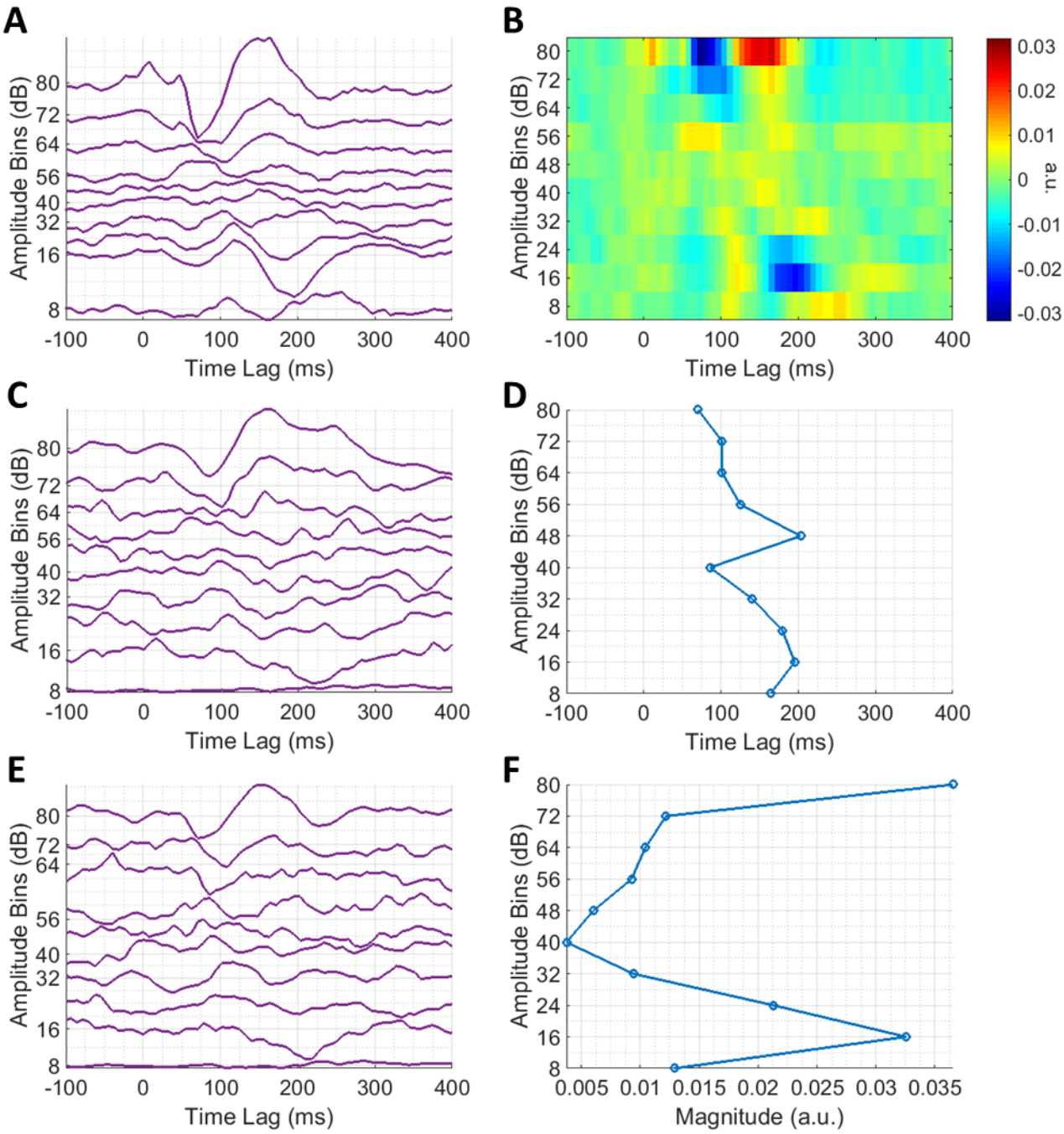

Figure 6.6: Analysis of Amplitude-Dependent Changes for the Speech Dataset.

A *– Group average AB envelope TRF.* ***B*** *– Image plot of group average AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***C*** *– Single subject AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***D*** *– N1 peak latencies across group average AB envelope TRF bins.* ***E*** *– Single subject AB envelope TRF, plotted to minimise the difference between adjacent traces.* ***F*** *– P1-N1 peak-to-peak amplitudes across group average AB envelope TRF bins. Please note, these findings represent amplitude-dependent changes at a single representative channel over left central scalp for the speech dataset.*

outperformed the individual onset envelope and AB envelope models, for both the AM BBN (see Table 6.3) and speech datasets (see Table 6.4), suggesting that they are capturing complementary information on envelope tracking in the cortex (Figure 6.7A and B). These results were also seen when comparing the models using AIC (all p <0.001; Wilcoxon signed rank test).

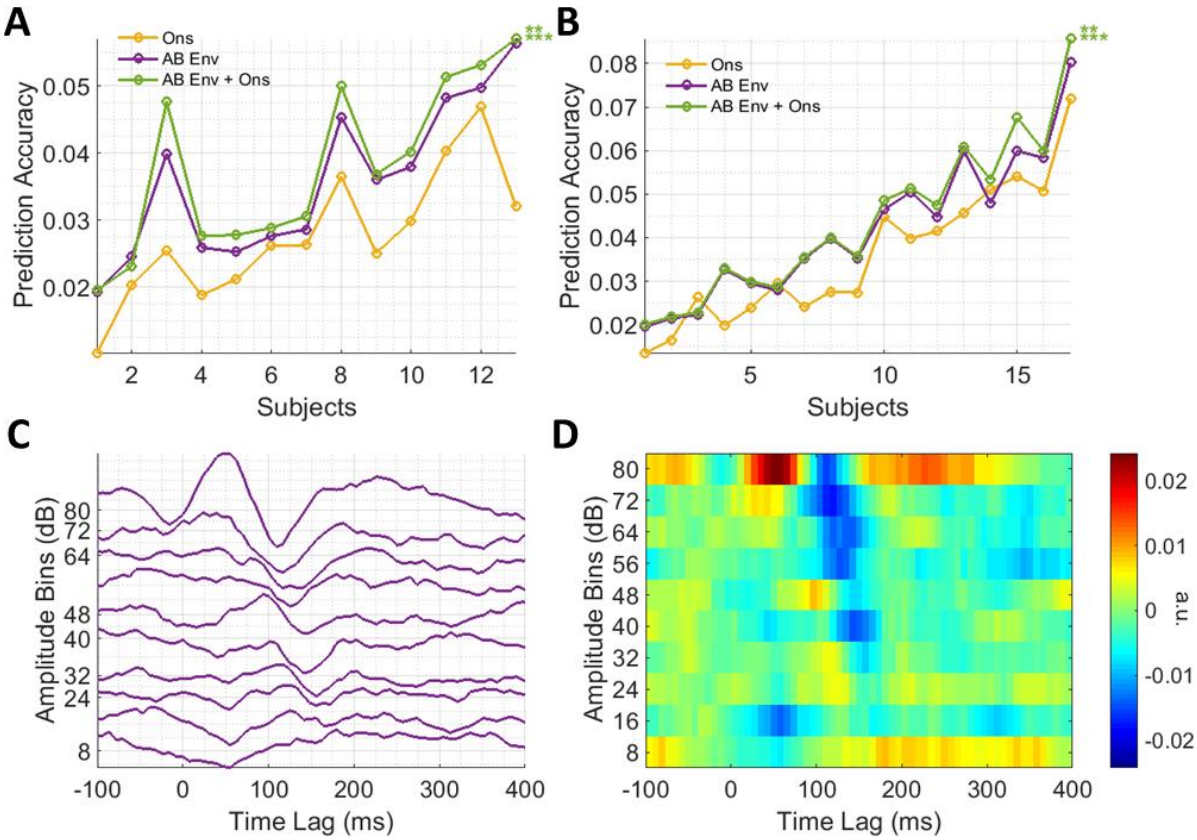

Figure 6.7: Prediction Accuracies and Analysis of Amplitude-Dependent Changes for both the AM BBN and Speech Datasets

A *– Prediction accuracies for each model and subject for the AM BBN dataset.* ***B*** *– Prediction accuracies for each model and subject for the speech dataset.* ***C*** *– Group average AB onset envelope TRF for the AM BBN dataset, plotted to minimise the difference between adjacent traces.* ***D –*** *Image plot of group average AB onset envelope TRF for the AM BBN dataset.*

One obvious extension to this approach then might be to also amplitude bin the onset envelope representation, producing an AB onset envelope model. However, while this AB onset envelope TRF exhibits a similar dependence on stimulus amplitude to the AB envelope TRF (Figure 6.7C and D) and significantly outperformed the onset envelope model alone for the AM BBN dataset ($t(12) = 3.887$, $p < 0.05$, $d = 1.078$) although not for the speech dataset ($t(16)$, $p > 0.05$), the combined AB envelope plus AB onset envelope model failed to

outperform the combined AB envelope plus onset envelope model for either the AM BBN or speech datasets.

Table 6.3: Comparison Between AB Envelope plus Onset Envelope and Other Models for the AM BBN Dataset

Model	df	t	p	d
Onset Envelope	12	5.717	< 0.001	1.586
AB Envelope	12	4.184	< 0.01	1.161

Table 6.4: Comparison Between AB Envelope plus Onset Envelope and Other Models for the Speech Dataset

Model	df	t	p	d
Onset Envelope	16	6.139	< 0.001	1.489
AB Envelope	16	3.312	< 0.01	0.8032

6.4 Discussion

Despite it long being known that the latency and morphology—and not just the magnitude—of auditory system responses are dependent on the stimulus amplitude, this has been overlooked in previous efforts at linearly modelling the auditory system. Here we have shown that by allowing the stimulus-response model to vary as a function of the stimulus amplitude, we can improve the modelling of responses to continuous auditory stimuli.

Specifically, we saw that by amplitude binning the envelope and then using that to fit a multivariate TRF, we could improve the prediction accuracy over the standard envelope model with a very large effect size for both the AM BBN and speech datasets. This was not the case for the SPL envelope or onset envelope models however, which both failed to outperform the standard envelope model. We also evaluated the offset envelope—created by half-wave rectifying the negative portion of the first-derivative of the envelope and then using that to fit a univariate TRF—and derivative envelope models, but again, neither managed to outperform the envelope model for either dataset and indeed mostly performed worse. Finally, we saw that by combining the AB envelope and onset envelope models, we could further improve the prediction accuracy over the AB envelope model with a large effect size for both the AM BBN and speech datasets.

Interestingly, despite having lower prediction accuracies overall, the improvement in prediction accuracy was greater for the AM BBN dataset. This is likely due to the differences in amplitude distribution seen between the two types of stimuli (Figure 6.1E and F). While the

speech stimuli predominantly vary within a narrow amplitude range, the wider active amplitude range of the AM BBN stimulus may allow it to benefit more from taking amplitude-dependent variations into account. The reason that the prediction accuracies were higher for the speech dataset overall is likely due to attention effects, e.g., as seen in O'Sullivan et al., 2014—albeit it in that case with two competing speech streams and reconstruction accuracy.

Previous work has shown that the use of other stimulus representations can also improve modelling performance. For example, for speech it has been shown that models based on spectrograms, phonemes, and phonetic features, outperform those based on the standard envelope (Di Liberto et al., 2015). However, for each of these stimulus representations, the same assumption of unchanging TRF morphology applies. For categorical representations such as those reflecting the phonemic/phonetic content of the speech, this could be considered a strength, but for lower-level representations such as the spectrogram, this could be considered a weakness.

While 80 and 72 minutes of data were used for the AM BBN and speech analyses respectively, less data would likely have sufficed. As with TDA, the inclusion of more and more data in a TRF analysis tends to come with diminishing returns. However, as this was an exploratory study, we wanted to ensure that we used enough data to test our hypothesis, and not leave its validation potentially inconclusive on account of insufficient SNR. That is not to say that the amount of data we used was excessive either. As can be seen in Figure 6.4C and E, and Figure 6.6C and E, the SNRs of some of the single-subject TRFs could still stand to be improved. Now that we have shown that this approach works, future work could focus on determining the minimum amount of data, number of channels, etc., that would be needed to efficiently conduct an analysis of this type—a crucial step if such methods are to be used in clinical environments.

In summary, here we have shown that by allowing the stimulus-response model to vary as a function of the stimulus amplitude, we can improve the modelling of responses to continuous auditory stimuli, and that the inclusion of an onset stimulus representation can improve this performance even further. This obviously has implications for how people model auditory processing in humans, but, more generally, points to the importance of incorporating stimulus-dependencies when modelling the activity of sensory systems.

Chapter 7. General Discussion

7.1 Introduction

In this thesis, a number of novel approaches for deriving TRFs have been put forward, and the merits and limitations of each discussed in-line. In this chapter, we will discuss this work in the broader context of the field, with a particular focus on the exciting potential for these approaches to be combined. As previously mentioned, the work presented in this thesis represents initial exploratory attempts at developing new approaches for studying and assessing the human auditory system. Future work—beyond improving upon these approaches—could focus on determining the minimum amount of data, number of channels, etc., that would be needed to efficiently utilise these approaches in clinical environments.

7.2 Low-Level Assessment

In Chapter 4 we introduced the VSFR approach which—with appropriate scaling and regularisation—enabled us to derive robust multiple-latency responses from the human auditory system. While this approach improves upon previous efforts at recovering multiple-latency responses from the human auditory system—particularly those involving the use of discrete stimuli—it does come with additional complexity in terms of the tuning of analysis parameters, e.g., those used to generate the sigmoid that determines the varying sampling rate across time-lag. Future work could perhaps focus on the use of objective methods for determining optimal parameter values, e.g., as is currently done when tuning the lambda value, i.e., cross-validation (see Chapter 3 for details).

We also showed in Chapter 4 that it was possible to derive narrowband ABRs reminiscent of derived-band ABRs (Don and Eggermont, 1978), using narrowband envelopes

recovered from the AM BBN stimulus. Theoretically then, if we were to combine these two approaches, i.e., to create a FB VSFR approach, we should be able to recover narrowband multiple-latency responses. This could potentially provide insight into not only the frequency, but also the neural source of any deficit—something that would typically require three separate analyses using other approaches. AM BBN is likely not the optimal stimulus for this purpose, however, and stimuli with non-flat spectra or even multiple narrowband stimuli—as was used in Chapter 5—may be preferred.

Such an approach could also potentially be useful for determining hearing thresholds. As we saw in Chapter 5, backward modelling—while typically displaying greater sensitivity than forward modelling—has its limitations—not least of which the lack of neurophysiological interpretability—and it was suggested that forward modelling and response detection might provide a useful alternative. This is particularly pertinent given the latter's already successful use with other objective approaches, e.g., AEP, ASSR, etc. If such an approach were to prove successful, it could also potentially allow more naturalistic stimuli such as speech or music to be used to assess hearing function—something that is not currently possible with other approaches. Such assessments could then potentially be conducted in real-time to account for environmental changes, e.g., going from a taxi to a live music venue, or perhaps negate the need for formal testing sessions altogether. It is not clear that the moving threshold approach would be the optimal method to exploit these new data, however, and the development of alternative approaches may be required.

One such approach could be to combine the FB VSFR approach with the AB approach, thus forming the FAB, i.e., frequency- and amplitude-binned, VSFR approach. Specifically, by amplitude binning the stimulus feature in each frequency bin and then using it to fit a multivariate TRF. Beyond likely further improving the prediction accuracies of the associated models beyond what was already achieved in Chapter 6, this could theoretically also provide frequency- and level-specific measures of auditory processing from which inferences of hearing function could be made, e.g., the presence or lack of a discernible peak at a particular frequency and level could help to determine a subject's hearing threshold. It is important to note however that this representation would be high-dimensional and so would come with increased computational requirements as well as an increased chance of overfitting.

As already alluded to, this approach would not be restricted to threshold detection and could potentially also provide other measures of auditory processing. For example, stimulus-dependent changes in ASSR amplitudes have been used as a measure of behavioural loudness

growth, i.e., the relationship between sound intensity and perceived loudness (Marks and Florentine, 2011; Van Eeckhoutte et al., 2016, 2018). Changes in TRF amplitude across amplitude bins could perhaps be used in a similar manner.

7.3 Low- and High-Level Assessment

Of course, this type of approach is not restricted solely to the measure of low-level processing. Indeed much of the work that has been done using TRF estimation in recent years has focused almost exclusively on high-level processing, e.g., while studying the so-called cocktail party attention problem (O'Sullivan et al., 2014), as well as linguistic processing at the level of phonemes (Di Liberto et al., 2015) and semantics (Broderick et al., 2018). As already mentioned, these are not entirely separate endeavours: low-level processing is an essential component of any higher-level processing, and so studying the interaction between these levels is important.

One approach for doing this could be to add interaction terms to the TRF, as suggested by Maddox and Lee (2018). This could be useful for example for studying the interaction between parietal α power and ABR amplitude (Maddox and Lee, 2018). This approach could perhaps be extended by replacing the speech envelope with the phonetic feature plus spectrogram representation as suggested by (Di Liberto et al., 2015), thus incorporating more precise measures of both low- and high-level processing. This could perhaps be further extended by amplitude binning the spectrogram so that even more of the variance would be accounted for, whilst also making other measures of low-level speech processing available. This could be important given that certain high-level processes such as selective attention have been shown to correlate with physiological measures of suprathreshold encoding in subcortical regions (Ruggles et al., 2011).

7.4 Decoding

As we saw in Chapter 6, researchers interested in improving the performance of their envelope tracking measures could benefit from using the AB—and potentially FAB VSFR—approach and/or including the onset envelope as part of their stimulus representation. The sensitivity and robustness of such measures could be further improved, however, if this work was adapted into a "decoding" framework. Such approaches have become quite popular in recent years and often involve mapping backward from the multivariate neural data to reconstruct an estimate of the

univariate feature that caused those data, i.e., backward modelling. As mentioned, this approach takes advantage of the large increase in modelling performance that comes with incorporating all the neural data simultaneously into one multivariate mapping. As such, it would be practically valuable to incorporate the FAB VSFR approach into a multivariate-to-multivariate decoding framework. Such frameworks have been implemented before for multivariate auditory stimuli, e.g., Mesgarani et al. (2009), and there are a number of flexible methods available that would be well suited to such a task, e.g., de Cheveigné et al. (2018).

If processing power and/or memory is at a premium, however, e.g., in the case of a smart-hearing aid that can "steer" its microphones based on the listener's attention, an alternative approach might be to use the FAB VSFR approach to create a frequency- and amplitude-compensated univariate feature, e.g., amplitude envelope. This is similar in principle to the gammachirp compensation that was applied to the AM BBN envelope in Chapter 4, except that the compensation would be subject-specific rather than being based on normative data. It could be calculated once using the FAB VSFR approach, and then applied directly to the univariate envelope thereafter, thus providing most of the benefit of the FAB VSFR approach while minimising the demands on hardware.

7.5 Summary and Conclusions

The body of work in this thesis provides a novel approach for efficiently deriving low-level multiple-latency responses, an initial attempt at a novel approach for objectively determining hearing thresholds, and a number of neurophysiologically-inspired stimulus representations that can enhance response derivation. Suggestions for improvements upon each approach have been outlined, as have the potential benefits of combining these approaches, not just with each other, but also with the large body of work that has already been done, studying high-level auditory processing. Such efforts could help to foster a better understanding of human auditory processing and related deficits, more diagnostically useful objective measures of hearing function along the auditory pathway, and ultimately, hopefully, better interventions.

Bibliography

Aiken SJ, Picton TW (2008) Human Cortical Responses to the Speech Envelope: Ear Hear 29:139–157.

Bardy F, Dillon H, Van Dun B (2014) Least-squares deconvolution of evoked potentials and sequence optimization for multiple stimuli under low-jitter conditions. Clin Neurophysiol 125:727–737.

Bardy F, Van Dun B, Dillon H (2015) Bigger Is Better: Increasing Cortical Auditory Response Amplitude Via Stimulus Spectral Complexity. Ear Hear 36:677–687.

Beagley HA, Knight JJ (1967) Changes in Auditory Evoked Response with Intensity. J Laryngol Otol 81:861–873.

Bear MF, Connors BW, Paradiso MA (2007) Neuroscience. Lippincott Williams & Wilkins.

Bidelman GM (2015) Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. J Neurosci Methods 241:94–100.

Bieser A, Müller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp Brain Res 108:273–284.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Broderick MP, Anderson AJ, Liberto GMD, Crosse MJ, Lalor EC (2018) Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Curr Biol 28:803-809.e3.

BSA (2011) Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. Read Br Soc Audiol.

Bunch CC (1922) Measurement of acuity of hearing throughout the tonal range. Psychol Monogr 31:45–82.

Bunch CC (1923) New Procedure in tests of auditory acuity. JOSA 7:1161–1166.

Burkard R, Shi Y, Hecox KE (1990) A comparison of maximum length and Legendre sequences for the derivation of brain-stem auditory-evoked responses at rapid rates of stimulation. J Acoust Soc Am 87:1656–1664.

Burkard RF, Eggermont JJ, Don M (2007) Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins.

Campbell KB, Picton TW, Wolfe RG, Maru J, Baribeau-Braun J, Braun C (1981) Auditory potentials. Sensus 1:21–31.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Netw Comput Neural Syst 12:199–213.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci 10.

Dau T, Wegner O, Mellert V, Kollmeier B (2000) Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. J Acoust Soc Am 107:1530–1540.

Davis H, Hirsh SK (1979) A Slow Brain Stem Response for Low-Frequency Audiometry. Audiology 18:445–461.

de Cheveigné A, Parra LC (2014) Joint decorrelation, a versatile tool for multichannel data analysis. NeuroImage 98:487–505.

de Cheveigné A, Wong DDE, Di Liberto GM, Hjortkjær J, Slaney M, Lalor E (2018) Decoding the auditory brain with canonical component analysis. NeuroImage 172:206–216.

Delgado RE, Özdamar Ö (2004) Deconvolution of evoked responses obtained at high stimulus rates. J Acoust Soc Am 115:1242–1251.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol 25:2457–2465.

Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Don M, Eggermont JJ (1978) Analysis of the click-evoked brainstem potentials in man using high-pass noise masking. J Acoust Soc Am 63:1084–1092.

Don M, Eggermont JJ, Brackmann DE (1979) Reconstruction of the Audiogram Using Brain Stem Responses and High-Pass Noise Masking. Ann Otol Rhinol Laryngol 88:1–20.

Don M, Masuda A, Nelson R, Brackmann D (1997) Successful detection of small acoustic tumors using the stacked derived-band auditory brain stem response amplitude. Am J Otol 18:608–21.

Downs MP, Doster ME, Weaver M (1966) Dilemmas in Identification Audiometry. J Occup Environ Med 8:656.

Drennan DP, Lalor EC (2019) Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. eNeuro:ENEURO.0082-19.2019.

Elberling C, Cebulla M, Stürzebecher E (2007a) Simultaneous multiple stimulation of the auditory steady-state response (ASSR). Proc Int Symp Audit Audiol Res 1:201–210.

Elberling C, Don M (2010) A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses. J Acoust Soc Am 128:2955–2964.

Elberling C, Don M, Cebulla M, Stürzebecher E (2007b) Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay. J Acoust Soc Am 122:2772.

Elberling C, Kristensen SGB, Don M (2012) Auditory brainstem responses to chirps delivered by different insert earphones. J Acoust Soc Am 131:2091–2100.

Eysholdt U, Schreiner C (1982) Maximum Length Sequences - A Fast Method for Measuring Brain-Stem-Evoked Responses. Audiology 21:242–250.

Fiedler L, Obleser J, Lunner T, Graversen C (2016) Ear-EEG allows extraction of neural responses in challenging listening scenarios; A future technology for hearing aids? In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 5697–5700.

Fiedler L, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, Obleser J (2017) Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J Neural Eng 14:036020.

Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci 78:2643–2647.

Geisler CD, Frishkopf LS, Rosenblith WA (1958) Extracranial Responses to Acoustic Clicks in Man. Science 128:1210–1211.

Goldstein R, Rodman LB (1967) Early Components of Averaged Evoked Responses to Rapidly Repeated Auditory Stimuli. J Speech Lang Hear Res 10:697–705.

Gonçalves NR, Whelan R, Foxe JJ, Lalor EC (2014) Towards obtaining spatiotemporally precise responses to continuous sensory stimuli in humans: a general linear modeling approach to EEG. Neuroimage 97:196–205.

Hall JW (1992) Handbook of Auditory Evoked Responses. Allyn and Bacon.

Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, Bießmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87:96–110.

Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H (2012) Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49:322–334.

Holt F, Özdamar Ö (2016) Effects of rate (0.3–40/s) on simultaneously recorded auditory brainstem, middle and late responses using deconvolution. Clin Neurophysiol 127:1589–1602.

Irino T, Patterson RD (2006) A Dynamic Compressive Gammachirp Auditory Filterbank. IEEE Trans Audio Speech Lang Process 14:2222–2232.

ISO 389-2 (1994) Acoustics - Reference zero for the calibration of audiometric equipment - Part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones.

ISO 389-6 (2007) Acoustics - Reference zero for the calibration of audiometric equipment - Part 6: Reference threshold of hearing for test signals of short duration.

Jewett DL, Caplovitz G, Baird B, Trumpis M, Olson MP, Larson-Prior LJ (2004) The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses. Clin Neurophysiol 115:2754–2775.

Jewett DL, Williston JS (1971) Auditory-evoked far fields averaged from the scalp of humans. Brain 94:681–696.

Jiang ZD, Brosi DM, Shao XM, Wilkinson AR (2000) Maximum Length Sequence Brainstem Auditory Evoked Responses in Term Neonates Who Have Perinatal Hypoxia-Ischemia. Pediatr Res 48:639–645.

Kohl MC, Strauss DJ (2016) A compact representation for the auditory full-range response and its fast denoising using an image filter based on the Radon Transform. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 5877–5880. Orlando, FL, USA: IEEE.

Kujawa SG, Liberman MC (2009) Adding Insult to Injury: Cochlear Nerve Degeneration after “Temporary” Noise-Induced Hearing Loss. J Neurosci 29:14077–14085.

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur J Neurosci 31:189–193.

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. J Neurophysiol 102:349–359.

Lasky RE (1997) Rate and adaptation effects on the auditory evoked brainstem response in human newborns and adults. Hear Res 111:165–176.

Lehmann D (1987) Principles of spatial analysis. Handb Electroencephalogr Clin Neurophysiol Methods Anal Brain Electr Magn Signals 1:309–354.

Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal Envelope Processing in the Human Left and Right Auditory Cortices. Cereb Cortex 14:731–740.

Lins OG, Picton TW (1995) Auditory steady-state responses to multiple simultaneous stimuli. Electroencephalogr Clin Neurophysiol Potentials Sect 96:420–432.

Lü Z-L, Williamson SJ, Kaufman L (1992) Human auditory primary and association cortex have differing lifetimes for activation traces. Brain Res 572:236–241.

Maddox RK, Lee AKC (2018) Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 5:ENEURO.0441-17.2018.

Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) Independent Component Analysis of Electroencephalographic Data. In: Advances in Neural Information Processing Systems 8 (Touretzky DS, Mozer MC, Hasselmo ME, eds), pp 145–151. MIT Press.

Marks LE, Florentine M (2011) Measurement of loudness, part I: Methods, problems, and pitfalls. In: Loudness, pp 17–56. Springer.

Mesgarani N, David SV, Fritz JB, Shamma SA (2009) Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. J Neurophysiol 102:3329–3339.

Musiek FE, Baran JA, Pinheiro ML (1994) Introduction to Case Studies in Neuroaudiology. Neuroaudiology-Case Stud Ed Singul Publ Group.

Musiek FE, Chermak GD, Bamiou D-E, Shinn J (2018) CAPD: The Most Common ‘Hidden Hearing Loss.’ ASHA Lead.

Musiek FE, Geurkink NA (1981) Auditory Brainstem and Middle Latency Evoked Response Sensitivity near Threshold. Ann Otol Rhinol Laryngol 90:236–240.

O’Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex.

Pickles JO (2008) An Introduction to the Physiology of Hearing, 3 edition. London: Academic Press.

Picton TW (2010) Human Auditory Evoked Potentials. Plural Publishing.

Picton TW (2013) Hearing in Time: Evoked Potential Studies of Temporal Processing. Ear Hear 34:385–401.

Picton TW, Ouellette J, Hamel G, Smith AD (1979) Brainstem evoked potentials to tonepips in notched noise. J Otolaryngol 8:289–314.

Power AJ, Reilly RB, Lalor EC (2011) Comparing linear and quadratic models of the human auditory system using EEG. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, pp 4171–4174.

Ross B, Lütkenhöner B, Pantev C, Hoke M (1999) Frequency-Specific Threshold Determination with the CERAgram Method: Basic Principle and Retrospective Evaluation of Data. Audiol Neurotol 4:12–27.

Ruggles D, Bharadwaj H, Shinn-Cunningham BG (2011) Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. Proc Natl Acad Sci 108:15516–15521.

Sams M, Hari R, Rif J, Knuutila J (1993) The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. J Cogn Neurosci 5:363–370.

Skoe E, Kraus N (2010) Auditory brainstem response to complex sounds: a tutorial. Ear Hear 31:302–324.

Srinivasan R, Nunez PL, Tucker DM, Silberstein RB, Cadusch PJ (1996) Spatial sampling and filtering of EEG with spline Laplacians to estimate cortical potentials. Brain Topogr 8:355–366.

Tanaka H, Komatsuzaki A, Hentona H (1996) Usefulness of Auditory Brainstem Responses at High Stimulus Rates in the Diagnosis of Acoustic Neuroma. ORL 58:224–228.

Tikhonov AN (1963) Solution of incorrectly formulated problems and the regularization method. Sov Math 4:1035–1038.

Tyberghein J, Forrez G (1971) Objective (E.R.A.) and Subjective (C.O.R.) Audiometry in the Infant. Acta Otolaryngol (Stockh) 71:249–252.

Valderrama JT, Alvarez I, Torre A de la, Segura JC, Sainz M, Vargas JL (2012) Recording of auditory brainstem response at high stimulation rates using randomized stimulation and averaging. J Acoust Soc Am 132:3856–3865.

Van Eeckhoutte M, Wouters J, Francart T (2016) Auditory steady-state responses as neural correlates of loudness growth. Hear Res 342:58–68.

Van Eeckhoutte M, Wouters J, Francart T (2018) Electrically-evoked auditory steady-state responses as neural correlates of loudness growth in cochlear implant users. Hear Res 358:22–29.

Van Maanen A, Stapells DR (2005) Comparison of multiple auditory steady-state responses (80 versus 40 Hz) and slow cortical potentials for threshold estimation in hearing-impaired adults. Int J Audiol 44:613–624.

Wang T, Zhan C, Yan G, Bohórquez J, Özdamar Ö (2013) A preliminary investigation of the deconvolution of auditory evoked potentials using a session jittering paradigm. J Neural Eng 10:026023.

WHO (2018) Deafness and hearing loss. World Health Organ Available at: http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [Accessed May 10, 2018].

Williams HL, Tepas DI, Morlock HC (1962) Evoked Responses to Clicks and Electroencephalographic Stages of Sleep in Man. Science 138:685–686.

Woldorff MG (1993) Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. Psychophysiology 30:98–119.

Wong DDE, Fuglsang SA, Hjortkjær J, Ceolini E, Slaney M, de Cheveigné A (2018) A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. Front Neurosci 12 Available at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00531/full.

Wu MC-K, David SV, Gallant JL (2006) Complete Functional Characterization of Sensory Neurons by System Identification. Annu Rev Neurosci 29:477–505.