University of Southampton

Faculty of Engineering and Physical Sciences

Institute of Sound and Vibration Research

Exploring Factors Important for Clinical Application of Cortical Responses to Continuous Speech

by

Ghadah Aljarboa

ORCID ID [https://orcid.org/0000-0003-3442-9211]

Thesis for the degree of Doctor of Philosophy

February 2025

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences

Institute of Sound and Vibration Research

Thesis for the degree of Doctor of Philosophy

Exploring Factors Important for Clinical Application of Cortical Responses to Continuous Speech

by

Ghadah Aljarboa

There is considerable interest in neural responses to continuous speech. Techniques for analysing these responses typically involve tracking EEG change due to stimulus features, such as the amplitude envelope. However, the clinical utility of these measurements, especially for challenging to test subjects such as infants with hearing aids, remains under-explored. This thesis aimed to investigate the clinical feasibility of neural tracking as an objective test for aided speech detection in infants. This aim was tackled through four studies designed to test factors essential for future application in infant testing in clinical environments. These factors included the feasibility of detecting responses in single-channel EEG recordings, detection time, effects of stimulus intelligibility, and attention. The two approaches used to analyse EEG signals were the temporal response function (TRF) and cross-correlation.

The first study assessed the effectiveness of single-channel EEG testing, achieving a 100% detection rate using cross-correlation within a detection time appropriate for clinical application. The second study focused on speech intelligibility effects during passive listening in recordings of cortical responses via single-channel EEG. The responses to speech-modulated noise demonstrated greater robustness regarding detectability and detection times than natural speech, indicating the potential utility of non-language-specific stimuli. Nevertheless, detection rates fell below 100%, potentially due to passive listening or shorter recording durations compared to the first study. The third study evaluated the envelope distortion induced by hearing aids using various stimuli. It found that the envelope distortion from the International Speech Test Signal (ISTS) was similar to that of natural speech, in contrast to speech-modulated noise, which exhibited significantly lower envelope distortion. The fourth study investigated the impact of different distortion levels on response detection using ISTS recordings from the third study. Higher levels of envelope distortion significantly lowered detectability and increased detection times, though using the envelope measured at the hearing aid output for detection analysis significantly improved these metrics. Additionally, no impact of attention on response detectability was observed.

In conclusion, single-channel EEG analysis showed variable detectability across different stimuli and conditions, suggesting that signal processing methods and recording times may still need to be optimised. The ISTS stimulus produced results comparable to natural speech, supporting its potential clinical use as a non-language-specific option. However, detectability was compromised in aided condition with high levels of envelope distortion. Using the speech stimulus from the output of a hearing aid (as opposed to the input signal to the aid) shows potential for improving response detectability. Additionally, the study demonstrated that neural tracking can be recorded under passive listening conditions, which could be important when testing infants.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Ghadah Aljarboa

Title of thesis: Exploring Factors Important for Clinical Application of Cortical Responses to Continuous Speech

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

ALJARBOA, G. S., BELL, S. L. & SIMPSON, D. M. 2023. Detecting cortical responses to continuous running speech using EEG data from only one channel. International Journal of Audiology, 62, 199-208.

Signature: Date:19/02/2025

Acknowledgements

First and foremost, I would like to praise God for granting me blessings throughout my studies and for giving me the strength to complete this thesis.

I want to express my deepest gratitude to my primary supervisor, Steve Bell, for his continuous and unwavering support, guidance, and encouragement throughout my PhD journey. His invaluable expertise, constructive suggestions, and constant feedback have been instrumental in shaping my research and helping me navigate through the challenges of this project. I am genuinely grateful for his dedication and commitment to my success.

Additionally, I extend my heartfelt thanks to my co-supervisor, David Simpson, for his invaluable assistance with the technical aspects of my project, particularly in the area of signal processing. His clear explanations, insightful advice, and patient guidance were critical in helping me overcome complex challenges. Moreover, his continuous encouragement and positive attitude provided me with the confidence and motivation needed.

I would also like to extend my sincere gratitude to Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, and the Saudi Cultural Bureau for providing me with the scholarship and continuous financial support which made this research possible.

A special thanks goes to my colleagues in the Signal Processing Audio and Hearing Group (SPAR), especially Suwijak Deoisres, for allowing me to join his experiment and for his assistance in writing and editing some of the analysis codes. Your collaboration and support have been invaluable.

Last but certainly not least, my deepest thanks go to my family. To my parents, Jawaher AlZaid and Salem Aljarboa: I would not be where I am today without your unconditional love and unwavering support throughout my life. To my husband, Yaser AlMarzouq, thank you for being there whenever I needed emotional support or a boost of self-esteem. Your presence made the challenging PhD years so much more enjoyable.To my lovely daughters, Yara and Ghiada, thank you for being the joyful distractions, time managers, and sources of happiness I needed to keep moving forward. You brought balance and perspective to my life during this journey. Finally, to my siblings, Abdulmalik, Turky, and Sara, thank you for your support, encouragement, and for always being there when I needed it. Your presence in my life has been a constant source of strength.

Definitions and Abbreviations

ABR ..... Auditory brainstem responses

AEP..... Auditory evoked potential

ALR..... Auditory late response

AMLR ..... Auditory middle latency responses

ASSR..... Auditory steady-state responses

BKB..... Bamford-Kowal-Bench

BSA..... British Society of Audiology

cABR..... Complex - Auditory brainstem responses

CAEP ..... Cortical Auditory Evoked Potential

CI ..... Cochlear Implant

CR ..... Compression ratio

CT..... Compression threshold

ECochC ..... Electrocochleography

EDI ..... Envelope distortion index

EEG ..... Electroencephalography

FFR..... Frequency following response

HA ..... Hearing aid

LAeq ..... A-weighted equivalent continuous sound level

ISTS ..... International Speech Test Signal

MEG..... Magnetoencephalography

NR ..... Noise reduction

mTRF ..... Multivariate temporal response function

SIN ..... Speech in noise

SNR ..... Signal to noise ratio

SRT..... Speech Recognition Threshold

TRF ..... Temporal responses function

WRS..... Word recognition score

XCOR..... Cross-correlation

Chapter 1 Introduction

Access to sound is essential for humans to connect to and interact with their environment as well as others. Sounds that people are exposed to range from simple tones such as a doorbell to complex sounds such as speech or music. Given this variety, hearing loss can impact human life in diverse ways. According to the latest statistics from the World Health Organization (WHO), approximately 430 million people worldwide are affected by hearing loss, with about 34 million of these being children (World Health Organization, 2021). Unaddressed hearing loss in children can lead to delays in expressive language abilities and lower academic achievement (American Speech Language Hearing Association (ASHA), 2022). Early identification and intervention during infancy are therefore essential to mitigate the adverse effects of hearing loss.

Audiologists aim to provide patients with access to everyday sounds via interventions for hearing loss, such as hearing aids (HAs) or cochlear implants (CIs), in order to enhance patient quality of life. Early intervention, particularly for children with congenital hearing loss, is crucial for developing speech and language skills, which in turn can improve future academic performance (Ching, 2015, Meinzen-Derr et al., 2020). Following a diagnosis, assessing the benefits of HAs outside of clinical settings is critical during follow-up visits. However, there is currently no established objective clinical method to evaluate access to natural continuous speech sounds with HAs for infants or other groups of patients who cannot respond reliably to behavioural tests requiring attention and/or speech comprehension (Visram et al., 2023).

Speech perception outcome measures are vital for assessing how HA users perform in everyday listening situations. Most speech tests require behavioural responses, which may not be suitable for certain patient groups, such as infants. Infants require objective evaluation of speech perception because they have not yet developed the necessary motor and expressive language skills to respond to stimuli effectively. Similarly, they may not have the language skills that would allow them to demonstrate speech comprehension. One alternative approach is using auditory evoked potentials (AEPs) in response to speech stimuli, which do not require any behavioural responses (Hall, 2015).

Various studies have explored different approaches and analysis methods to assess AEPs in relation to speech stimuli. One such clinical tool, developed by the National Acoustic Laboratories (NAL), detects cortical responses to short speech stimuli (/m/, /g/, /t/) and is widely used in Australian clinics (Purdy and Kelly, 2001), whilst another approach is to investigate frequency-following responses (FFRs) to vowels (Skoe and Kraus, 2010). Most of those approaches rely on averaging the responses to repeated short stimuli. However, such

stimuli have limited environmental relevance and do not reflect how well HA and CI may work outside clinical settings.

What we hear daily encompasses a broad range of continuous and complex sounds, with speech being of primary concern in this work due to its range of temporal and spectral features and importance in infant development. Methods like those developed by the NAL group, which incorporate a variety of speech sounds covering a range of frequencies, represent an improvement over conventional tests that use clicks or tone bursts, yet they still cannot fully capture the range of natural speech. Recent advancements in computational capabilities have led to several new approaches for analysing EEG responses. These methods eliminate the need for averaging and enable the use of continuous speech as a stimulus by either directly applying the cross-correlation (XCOR) method (Aiken and Picton, 2008a) or modelling the relationship between stimulus and responses (Lalor et al., 2009). However, there are a number of questions that need to be answered before such methods can be used in clinics: (1) What is the detection sensitivity of these methods in individuals with normal hearing? (2) Can responses be recorded in reasonable timeframes? (3) Are methods sensitive to attention and if so, are they sensitive to the effects of HA processing on test stimuli?

1.1 Aim

The overarching goal of this project is to investigate the clinical feasibility of using continuous speech stimuli to measure aided cortical responses, with an eye toward future applications in assessing infant speech perception. This objective was pursued by examining the clinical applications and practicality of the current methods for measuring cortical responses to continuous speech. Whilst the long-term future aim is to develop methods that are applicable to infant testing, due to recruitment, ethical issues and the need to obtain well-controlled data, all studies were conducted on normal-hearing adults.

1.2 Original Contributions

1.2.1 Single channel EEG and individual detection to continuous speech:

An original contribution of this work is the demonstration of strong individual cortical detectability of continuous speech using a single-channel EEG. In the first study (Chapter 3), detection sensitivity reached 100% with intelligible speech as the stimulus. Furthermore, in the fourth study (Chapter 6), detectability remained high, at approximately 91%, when using the International Speech Test Signal (ISTS) as the stimulus.

1.2.2 The use of non-language-specific stimuli

This thesis investigated individual detection and detection time using non-language-specific continuous speech stimuli to evoke cortical responses—an area that, to the author's knowledge, has not been previously explored. The use of speech-modulated noise improved both detection rates and reduced detection time compared to natural speech. When comparing the ISTS detection results in Chapter 6 to the natural speech detection in Chapter 3, the outcomes were similar, with equivalent recording times.

1.2.3 Effect of attention on cortical responses to continuous speech

This thesis demonstrated that responses can be recorded in passive conditions with high detectability—approximately 91% using the ISTS stimulus, as shown in the fourth study (Chapter 6). However, with aided stimulation, detection rates decreased slightly with lower distortion and significantly with higher distortion.

1.2.4 Effect of hearing aid envelope distortion on cortical response detection

Unlike previous studies, the third study in this thesis quantified the amount of envelope distortion before measuring aided responses to continuous stimuli in the fourth study. The findings in Chapter 5 show that ISTS stimuli are distorted in a manner similar to natural speech when processed by an HA, whereas speech-modulated noise exhibited significantly lower distortion. Chapter 6 demonstrates the impact of envelope distortion on aided responses, evidenced by a significant reduction in detectability and an increase in mean detection time under higher distortion conditions compared to the unaided condition.

1.3 Publications

Conferences

- Aljarboa G. S., Bell S. L., and Simpson D. M. (2022) Comparing Cortical Responses to Continuous Speech and Speech Modulated Noise During Passive Listening. ‘Ear and Hear” Audiology and Audio conference. Southampton, UK. *Poster presentation*.
- Aljarboa G. S., Bell S. L., and Simpson D. M. (2023) Comparing Cortical Responses to Continuous Speech and Speech Modulated Noise During Passive Listening. International Evoked Response Audiometry Study Group. Cologne, Germany. *Poster presentation*.

Journal Articles

- Aljarboa G. S., Bell S. L., and Simpson D. M. (2022). Detecting cortical responses to continuous running speech using EEG data from only one channel. *International Journal of Audiology*, 1-10. *Published*.

- Aljarboa G. S., Bell S. L., and Simpson D. M. The detection of aided cortical responses to continuous speech-like stimuli in normal hearing adults, investigating the effect of envelope distortion. *In preparation for submission*.

1.4 Outline of thesis

Chapter 1 Introduction

This chapter serves as an introduction to the thesis, providing a concise overview of the significance of hearing in human communication and the methodologies employed to assess hearing in infants with hearing loss. It will also detail the motivation for this research, outline the purpose of the thesis, and highlight its contribution to the field of auditory evoked responses.

Chapter 2 Literature Review

This chapter provides a more in-depth background on the focus of the PhD project. It will begin with a comprehensive introduction to early detection of hearing loss in infants. It then discusses the current testing method used to assess speech perception through HAs and its limitations. Most of this section is dedicated to an extensive examination of various AEP tests found in the literature that utilise speech and related stimuli. This will lead to clearly identifying the gaps in the existing literature and the thesis's main objectives.

Chapter 3 Measuring Cortical Responses to Continuous Running Speech Using EEG Data From One Channel

This study explores the feasibility of measuring cortical responses to continuous running speech using EEG data from a single channel, focusing on comparing different detection approaches for speech-evoked responses. The methods of analysis include XCOR and the temporal response function (TRF), which were evaluated in terms of detection sensitivity and detection time.

Chapter 4 Comparing Cortical Responses to Continuous Speech and Speech-Modulated Noise During Passive Listening

This study assesses the effect of stimulus intelligibility in detecting cortical responses to continuous stimulus by comparing responses to continuous speech and speech-modulated noise during passive listening. The data in this study were jointly collected with Suwijak Deoisres, another PhD candidate.

Chapter 5 Quantifying the Effects of Hearing Aid Processing on the Stimulus Envelope

This study investigates the impact of HA processing on the stimulus envelope by measuring envelope distortion of three different speech stimuli: natural speech, speech-modulated noise, and the ISTS. The recordings obtained from this study served as stimuli for the subsequent EEG study.

Chapter 6 Investigating the Effect of Hearing Aid Envelope Distortion on Cortical Responses to Continuous Speech

This study evaluates the feasibility of detecting cortical responses to aided ISTS by measuring responses to stimuli prerecorded from HA output. It also examines the effect of attention by comparing passive and active listening conditions.

Chapter 7 General Discussion and Conclusion

This chapter combines the results of the different studies, discusses them, and positions them within the broader academic context. It also includes the project's limitations and suggestions for future work.

Chapter 2 Literature Review

This chapter provides background information on this research project and reviews related studies. It is divided into four main sections. The first section offers an overview of current clinical speech tests. The second section discusses various clinical AEP measurements. The third section explores using repeated short speech in AEP tests to assess speech perception. Finally, the fourth section explores AEP in response to running speech, including clinical application and feasibility.

2.1 Overview of current clinical speech tests

Early detection and intervention are key to overcoming the negative impacts of permanent hearing loss in children. Universal newborn screening programmes have been implemented worldwide to achieve this. For example, all newborns in the UK should be screened within four weeks of age. If they do not pass, they should be referred for complete audiological testing within four weeks (National Health Service (NHS), 2021). Following HA fitting (if required), follow-ups should be conducted every three months during the first and second years of life. Follow-up approaches should include real-ear measurements, unaided behavioural testing, functional aided testing (such as Ling 6 sound tests or speech-discrimination tasks) and age-appropriate questionnaires. Although behavioural speech tests are part of the guidelines, they are only suitable for older children with sufficient language ability to respond. To date, no objective clinical outcome measure of speech perception in the UK should be considered appropriate for infants fitted early with an aided hearing device (Visram et al., 2023).

Speech perception outcome measures provide insights into how HA users perform in everyday listening situations. These measures assess different stages of auditory skills, as described by Erber (1976): detection, discrimination, identification and comprehension. Current clinical practice to evaluate speech perception mainly involves behavioural testing, which requires a form of behavioural response from the participant. According to the British Society of Audiology (BSA), when assessing speech in adults, behavioural testing can be applied in a range of ways (British Society of Audiology, 2018). The primary method is speech threshold testing, such as the speech recognition threshold (SRT) approach, used to verify the accuracy of the patient's pure tone threshold and assess speech detection. The SRT is defined as the minimum speech level at which the subject can recognize 50% of the speech material presented (Van Tasell and Yanz, 1987). Word recognition score (WRS) is another standard speech test used to evaluate speech discrimination abilities at an audible level above the subject's threshold either in quiet

or in the presence of background noise. Both tests are limited in predicting aided speech perception in everyday communication (Taylor, 2007).

The BSA guidelines for testing speech perception in adults suggest using speech-in-noise (SIN) tests that include sentences as stimuli (BSA, 2018). Examples include the quick speech-in-noise test (QuickSIN) developed by Etymotic Research (ER, 2011), the Bamford-Kowal-Bench test (BKB) (Bench et al., 1979) and the City University of New York Sentences (CUNY) (Boothroyd et al., 1985). The QuickSIN test measures a subject's ability to understand SIN at different signal-to-noise ratio (SNR) levels. The BKB test was originally published to test children's speech recognition ability in quiet conditions (Bench et al., 1979) and later developed to test SIN (BKB-SIN) (Niquette et al., 2003). The BKB-SIN is similar to the QuickSIN in assessing speech recognition at different SNR levels using short, simple sentences (British Society of Audiology, 2018). The CUNY test provides a different measure by assessing speech comprehension at a fixed SNR with and without visual cues (e.g., lip-reading). The BSA guidelines recommend using SIN tests for baseline assessments and post-intervention performance evaluations.

Clinical behavioural speech tests can also test children, as long as age-appropriate norms are available for the tests. Speech detection tests such as the SRT can be applied with modified behavioural techniques, such as play audiometry or visual reinforcement audiometry (VRA) (McArdle and Hnath-Chisolm, 2015). Responses depend on the motor and language abilities of the child; examples include head turning, picture pointing and repeating words. The Ling 6 Sound Check tests different aspects of speech perception, including identification, detection and discrimination (Scollie et al., 2012). Children's speech recognition can also be examined using adult tests, such as BKB sentences and the BKB-SIN (McArdle and Hnath-Chisolm, 2015). Early Speech Perception Tests (ESPT) can test children as young as three (Geers and Moog, 2012). Speech testing for older children with hearing loss, especially cochlear implant users, is a common clinical practice (Eisenberg et al., 2010). However, speech tests are not generally carried out on infants and young children as they have not developed the language and motor abilities for behavioural responses, emphasising the need for an objective test.

Other measurements rely on the subjective evaluation of HAs and CIs. These can be self-reported questionnaires, such as the Abbreviated Profile Of Hearing Aid Benefit (APHAB) (Cox and Alexander, 1995), or parent-reported surveys, such as the Parents' Evaluation of Aural/Oral Performance of Children (PEACH) (Ching and Hill, 2007). However, these outcome measures might lack reliability as they rely solely on parental observations. Instead, an objective speech test using AEP response measurements would be more desirable in such cases to quantify accurately which speech sounds are or are not being represented in the brain, both with and without HA or other interventions. The following sections will discuss the general use of AEP and

its specific application as a speech outcome measurement tool for evaluating hearing loss interventions.

2.2 Auditory evoked potential measurements

AEPs are voltages that represent the neural activity produced by the auditory system in response to stimuli (Hall, 2015). They provide a means for monitoring changes in potential through transducers placed on the listener's head. This measure has many clinical and research applications in audiology. It is one of the main tools used in audiology clinics, mainly for assessing infants' hearing thresholds and for differential diagnosis. Auditory evoked responses are usually measured using electroencephalogram (EEG), magnetoencephalogram (MEG) or electrocochleography (ECochG). These objective technical processes measure auditory responses with high temporal resolution, making them appropriate for assessing auditory processing. ECochG measures the electrical potentials generated by the cochlea, typically through a minimally invasive procedure where an electrode is inserted (Gibson, 2017). These potentials can also be recorded non-invasively using an electrode placed in the external ear canal near the tympanic membrane (Attias et al., 2008). On the other hand, EEGs and MEGs are clinically more feasible as they can track neural processing at various levels of the auditory pathway, which is crucial for speech perception (Aiken and Picton, 2008b). MEGs are sensitive only to tangential-oriented components of cortical current sources, whereas the EEG is sensitive to both tangential and radial-oriented components of cortical current sources. This allows the EEG to measure a broader range of electromagnetic brain activity than the MEG (O'Sullivan et al., 2015). The EEG is highly popular in clinical and research settings due to its widespread availability and low cost. In contrast, MEG requires more complex equipment and must be conducted in specially shielded rooms.

The stimuli that evoke AEPs can range from short, repeated sounds to continuous speech. AEPs have very low amplitudes, typically ranging from 0.5 to 10 microvolts (μV) (Hall, 2015), compared to ongoing EEG activity, which ranges between 10 and 100 μV (Blinowska and Durka, 2006). This will result in a low SNR and makes detection challenging. The most well-known approach to overcome these issues is to average multiple responses to improve the SNR. This approach was built on the assumption that the responses are time-locked to the stimuli (Ruhm et al., 1967, Mendel and Goldstein, 1969). However, AEPs can vary in latency, which can affect the accuracy of the averaging outcome (Quiroga and Garcia, 2003).

Averaging requires a high number of repeated and identical stimuli, which makes it unsuitable for analysing continuous (running) speech. Although the averaging approach has proven to be very useful in estimating hearing sensitivity, it yields limited information about the human

auditory system's processing of stimulus temporal features (Lalor and Foxe, 2010). Using discrete and short stimuli to elicit responses does not reflect the human auditory system's abilities to analyse most sounds in real-life environments (Lalor et al., 2009). Recently, new approaches to analysing responses to continuous stimuli have been developed; some of these studies will be discussed in Section 2.4.

The results of an AEP test are usually interpreted visually by the examiner (e.g., an audiologist) based on normative data. Several alternative methods for objective and automated identification have been described in the literature. These methods typically use statistical parameters to identify the responses (Barajas, 1985, Lv et al., 2007, Chesnaye et al., 2018, Chesnaye et al., 2023). AEPs are usually classified depending on component latency and are represented in a time domain. The earliest responses are the ECochG and the auditory brainstem response (ABR), followed by the auditory middle latency response (AMLR) and the auditory late response (ALR) (Hall, 2015). Figure 2.1 illustrates the latencies of each of these responses. Auditory steady-state responses (ASSRs) are usually presented in the frequency domain (after performing Fourier Transforms) instead of the time domain since responses are expected at specific frequencies rather than specific time points (as is the case with ABRs, AMLRs or ALRs). Responses also vary based on the type of stimuli used to evoke them, which can range from short stimuli to continuous speech. The following section will discuss some primary responses used to assess hearing abilities in clinical audiology settings.

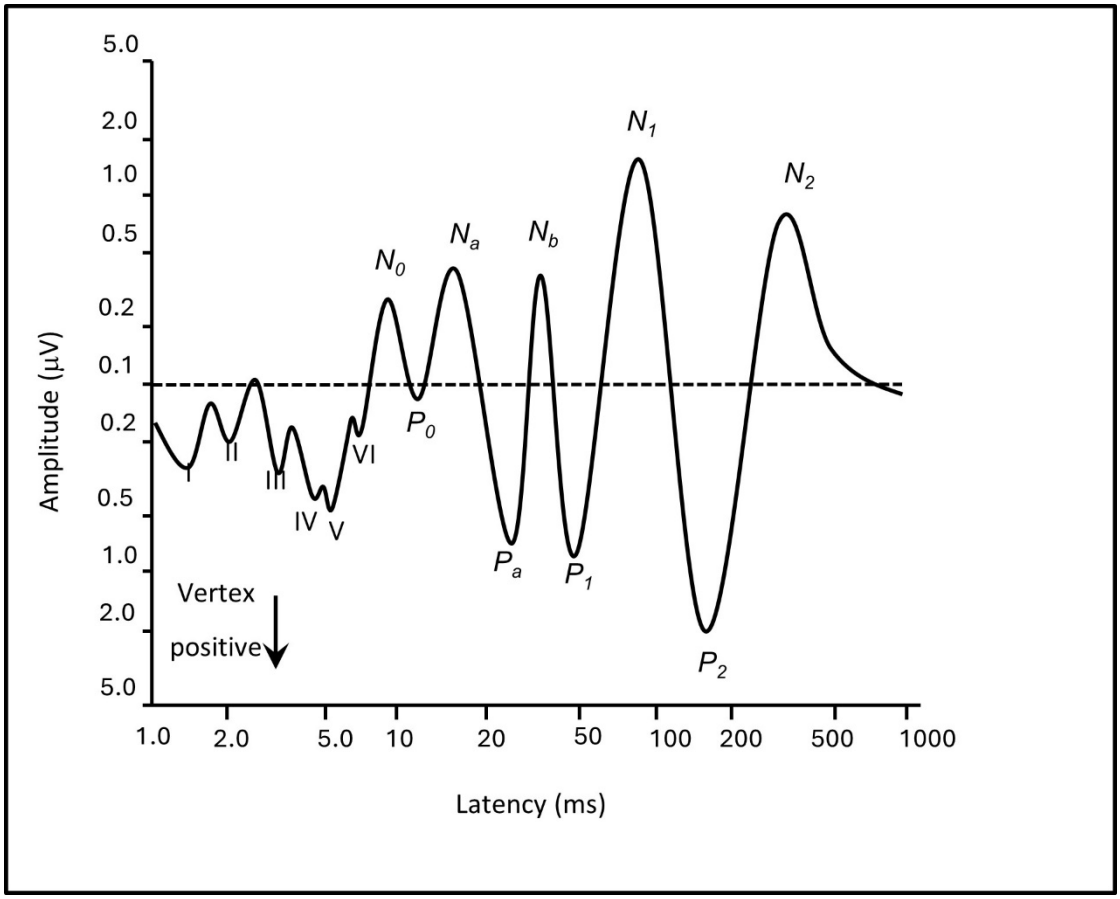

Figure 2.1 The AEP waveform showing ABR, AMLR, and ALR responses. *The figure illustrates the AEP waveform resulting from averaged responses to click stimuli, showing the ABR, AMLR and ALR. This illustration has been adapted from Pérez-González and Malmierca (2014), originally cited in Picton et al. (1974).*

2.2.1 Auditory brainstem responses

The ABR is the most popular and widely referenced type of response. ABRs are elicited for many clinical applications, including newborn hearing screening, infant hearing threshold estimation and differential diagnosis for the site of a lesion. Compared to late latency responses, the ABR is more clinically reliable and is not affected by the state of arousal (Hall, 2015). It can be evoked using a variety of stimuli, such as clicks, tone bursts, chirps (Hall, 2015) and short speech components (Skoe and Kraus, 2010). Responses are typically recorded via EEG using a one- or two-channel montage. After signal processing, the waveform of the response is presented in the time domain and commonly interpreted visually. The resulting waveform has five prominent peaks (I–V), where peak V is the most essential to detecting the presence of responses (see Figure 2.1). The lowest stimulus level at which wave V is detected twice determines the subject's hearing threshold.

Only a few articles in the literature investigate using click and tone burst ABR for aided testing (when a subject wears a HA). One such study suggests that aided ABR using short stimuli can cause significant distortion of the output with moderate to high gain non-linear HAs (Garnham et

al., 2000). Several other studies conclude that aided click and tone burst ABR does not provide a full picture of HA performance (Gorga et al., 1987, Brown et al., 1990). Brown et al. (1990) compared HA operation in response to continuous stimuli (noise and tones) and brief stimuli (tone-pips). The brief stimuli failed to activate the output-limiting circuitry, making them a poor representation of how continuous, real-world stimuli would be processed by HA and thus of aided hearing. When comparing ABR responses to tone-pips under aided and unaided conditions at the same intensity levels, no differences in the amplitude or latency of wave V were observed. The authors explained this finding as the adaptation of the ABR caused by the feedback from the HA and saturation from high levels of stimulation.

2.2.2 Auditory-steady state response

The ASSR is a brain response evoked by continuous stimuli, which can be frequency or amplitude-modulated (Dimitrijevic and Cone-Wesson, 2015). It is an essential part of the infant test battery, mainly for hearing threshold estimation, because it provides a frequency-specific estimation of the patient's hearing threshold. Responses are presented in the frequency domain, appearing as a peak in the spectrum corresponding to the modulation rate used. The detection and interpretation of the response are automated and verified statistically (Hall, 2015). However, clinical experience is required to assess the reliability and validity of the results and how they can be used to make a final diagnosis (Hall, 2015).

Several studies have investigated the use of ASSR as a HA outcome measure. The ASSR is considered more suitable for aided testing than the ABR because the continuous stimuli used better reflect HA processing than the rapid transient stimuli commonly used with evoked potentials (Dimitrijevic and Cone-Wesson, 2015). Using amplitudes and frequencies modulated to resemble human speech, Dimitrijevic et al. (2004) examined whether the ASSR correlates with the WRS in adults with hearing loss. A significant correlation was found between the WRS and the number of significant ASSR responses. The correlation value ranged between (r = 0.70 and r = 0.85) depending on stimulus frequency modulation. Moreover, the ASSR showed an increase in the number of significant responses under aided conditions compared to the unaided condition, indicating the potential applicability of such responses as an outcome measure of HA effectiveness. Although the study by Dimitrijevic et al. (2004) demonstrated a correlation between behavioural speech tests (WRS) and an objective speech testing tool (ASSR), it is not clear whether such an objective tool can be used independently to assess the benefits of hearing interventions or to determine the need for future interventions.

2.2.3 Auditory late responses

The ALR originates from the auditory cortex from roughly 50 ms post-stimuli (McArdle and Hnath-Chisolm, 2015). As with earlier responses, the ALR is evoked using repeated stimuli. The most clinically used ALR waves are the P1-N1-P2 complex, which ranges in latency between 50 ms and 200 ms (see Figure 2.1). Several studies have used the term 'cortical auditory evoked potential' (CAEP) to describe late responses. One difference between ALR and CAEP is that the latter is a general term, referring to all responses with cortical original. These include the AMLR and late responses including the P300 and mismatched negativity (MMN) (Hall, 2015). Unlike earlier responses, however, the ALR provides more information about sound processing up to a higher level.

The typical method for recording the CAEP is the 10-20 system, which organises 32 to 256 electrodes on the scalp (see Figure 2.2) (McArdle and Hnath-Chisolm, 2015). An advantage of this multichannel system is that it can identify the source of noises, such as eye blinks, which can profoundly affect the CAEP (McArdle and Hnath-Chisolm, 2015). The CAEPs are typically recorded from many electrodes over the scalp, especially with speech stimulation. However, electrodes located in the frontal portion of the scalp have the largest amplitude (Hall, 2015, McArdle and Hnath-Chisolm, 2015). In the case of repeated short stimuli, many studies have successfully detected CAEP using a single active channel with the electrode placed at the vertex position (Cz), the reference electrode positioned at the mastoid, and the ground electrode placed on the contralateral mastoid in adults and children (Lightfoot, 2016, Carter et al., 2010). Although a multichannel EEG setup can provide more insights into neural processing, single-channel analysis is more convenient for clinical practice, as it requires less expensive equipment and has a shorter setup time.

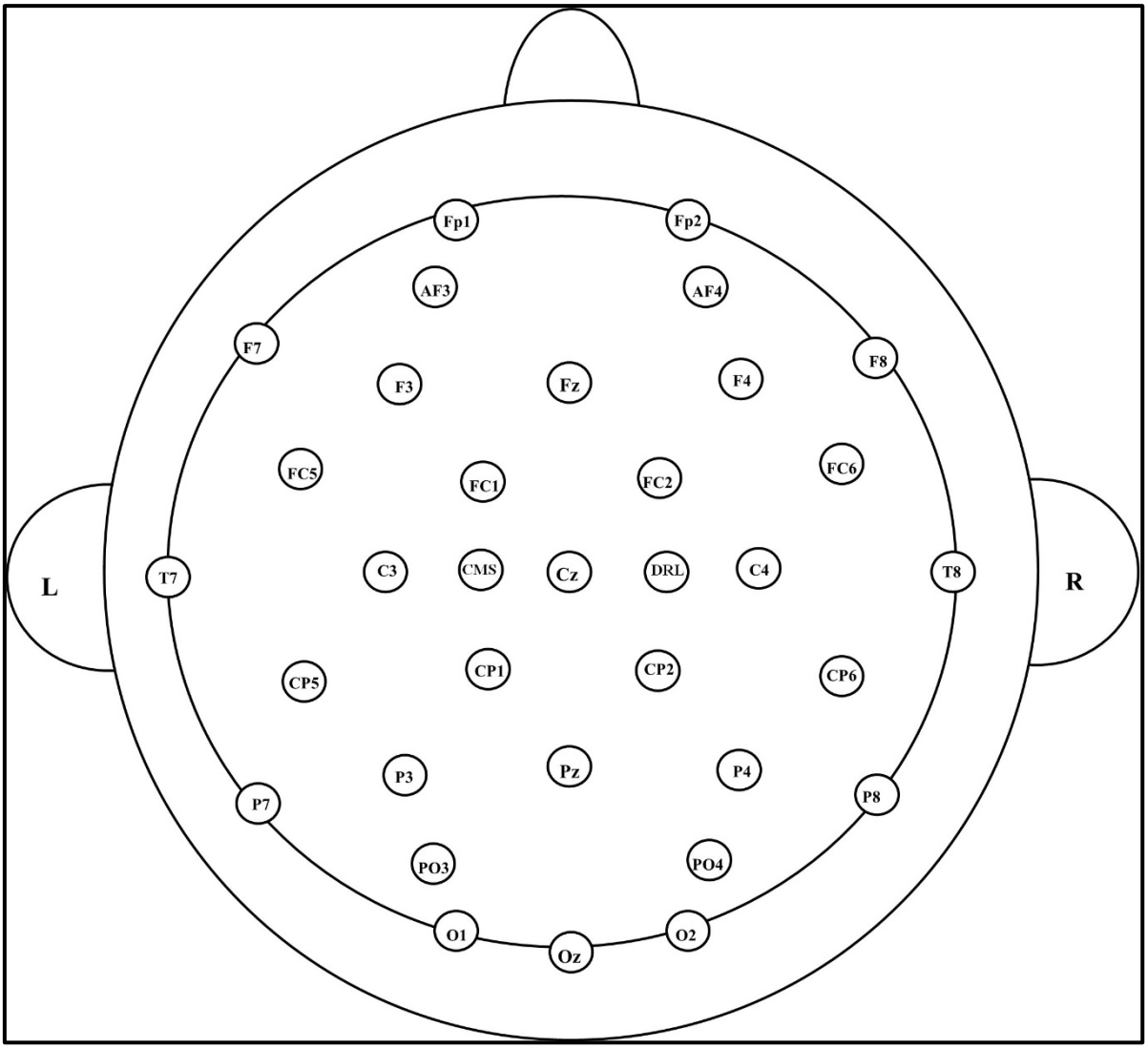

Figure 2.2 The BioSemi 32-channels cap. *The figure shows the electrode positioning of the BioSemi 32-channel cap (international 10–20 system of electrodes application), Redrawn from BioSemi (2015)*

The CAEP has recently gained clinical attention for HA evaluation, especially with speech stimuli. The NAL group, for example, uses the CAEP to measure HA performance in infants (Purdy and Kelly, 2001). The NAL's protocol will be reviewed in Section 2.3.3.

2.3 Auditory evoked potential for speech perception

As discussed earlier, behavioural or subjective tests typically assess speech perception abilities. Recently, studies have begun investigating the possibility of applying auditory evoked responses to evaluate speech perception. The primary rationale is to develop an objective clinical tool for assessing speech perception in quiet and noise conditions. This section will briefly discuss several approaches for using speech as a stimulus to evoke brain responses. The stimuli can range from simple short stimuli to continuous running speech, with numerous methods for analysing EEG responses.

2.3.1 Frequency following responses

The FFR is a broad term that refers to the measurement of synchronous brain responses to a stimulus originating from the subcortical level (Kraus et al., 2017). The FFR can provide more insight into sound processing in the brain by creating a close representation of the properties of the stimuli (Kraus et al., 2017). The amplitude of the FFR, like other brainstem responses, is relatively small compared to EEG activity, so it is detected using averages and then represented in the frequency domain (Choi et al., 2013). Different speech stimuli are commonly used to elicit the FFR, such as speech formants, vowels, speech fundamental frequencies or harmonics (Aiken and Picton, 2008a). The FFR represents both the fine structure and the envelope of speech stimuli. One approach to distinguish between spectral FFR and envelope FFR is to use alternating phase stimuli (Aiken and Picton, 2008a). The neural response to the envelope does not change with inverted stimuli, whereas the neural response to the fine structure does. Thus, the envelope FFR results from averaging responses to alternating polarity stimuli, whereas the spectral FFR results from subtracting responses to inverted stimuli from responses to original stimuli. Distinguishing spectral and envelope responses has been shown to provide more clinical information (Aiken and Picton, 2008a). While the envelope FFR ensures that important speech information reaches the nervous system, the spectral FFR only reveals significant responses to related harmonics (formants).

Most studies that use the envelope FFR use vowels as stimuli, either isolated or in words or sentences (Aiken and Picton, 2006, Choi et al., 2013, Vanheusden et al., 2019). Using vowels typically provides only limited information about speech bandwidth, limiting its suitability for aided measurements (Easwar et al., 2015b). Easwar et al. (2015a) efficiently measured the envelope FFR to naturally produced speech sounds with different acoustical features. They proposed a test paradigm to represent the bandwidth of speech stimuli and provide frequency-specific information. The envelope FFR was measured using 300 repeated sweeps of the speech token /susaʃi/, with each sweep including both polarities. The stimuli were low-pass filtered to 1 kHz, 2 kHz, and 4 kHz and presented at 50 dB and 65 dB SPL levels. They compared the effects of different stimuli bandwidths on the envelope FFR and the behavioural speech discrimination test score.

The findings demonstrate that when the cut-off frequency was raised, the number of detections and the summed amplitude of the envelope FFR increased. This increase showed a significant correlation with speech discrimination results. Moreover, a significant effect of stimulus level was observed, as increasing the stimulation level from 50 dB to 65 dB SPL significantly increased the response amplitude and, consequently, the number of detections. The results indicate that the envelop FFR elicited by /susaʃi/ is an effective way to assess speech

discrimination and has the potential to measure HA outcomes. The same proposed envelop FFR paradigm was tested in adults with hearing loss (Easwar et al., 2015c), indicating the envelope FFR's sensitivity to measuring aided responses. This method was tested further and showed a fair-to-good test-retest variability in normal-hearing adults (Easwar et al., 2020). Although this method shows promising results as an objective measure for verifying HA outcomes, the stimuli used are still some distance from being continuous natural speech.

2.3.2 Speech-ABR

One specific form of the FFR is the complex auditory brainstem response or cABR, which has been adopted to refer to responses to complex stimuli such as speech and music (Skoe and Kraus, 2010). The most studied stimulus used to evoke responses is a consonant-vowel (CV) syllables syllable, mainly the /da/ sound (Skoe and Kraus, 2010). As this type of stimulus consists of transient and sustained features, it will elicit onset and frequency-following responses, respectively. The FFR is the brainstem neural response to the periodicity of the vowel stimuli (Galbraith et al., 1995).

For cABR, the most common configuration is the single-channel montage recording (Skoe and Kraus, 2010). On average, about 6000 stimulus sweeps are required to evoke a detectable response. Responses are presented in the time domain, with latency represented by the onset of the stimuli. Amplitude is quantified by the root-mean-square magnitude of a specific region of the responses. Figure 2.3 shows the response onset's positive wave V, followed by the first negative wave A. The periodicity of the vowel is represented by the negative waves (D, E and F), with latency from 20 ms to 40 ms. Finally, the negative wave O shows the response offset (Chandrasekaran and Kraus, 2010). Other analyses are possible, but these parameters are the most common.

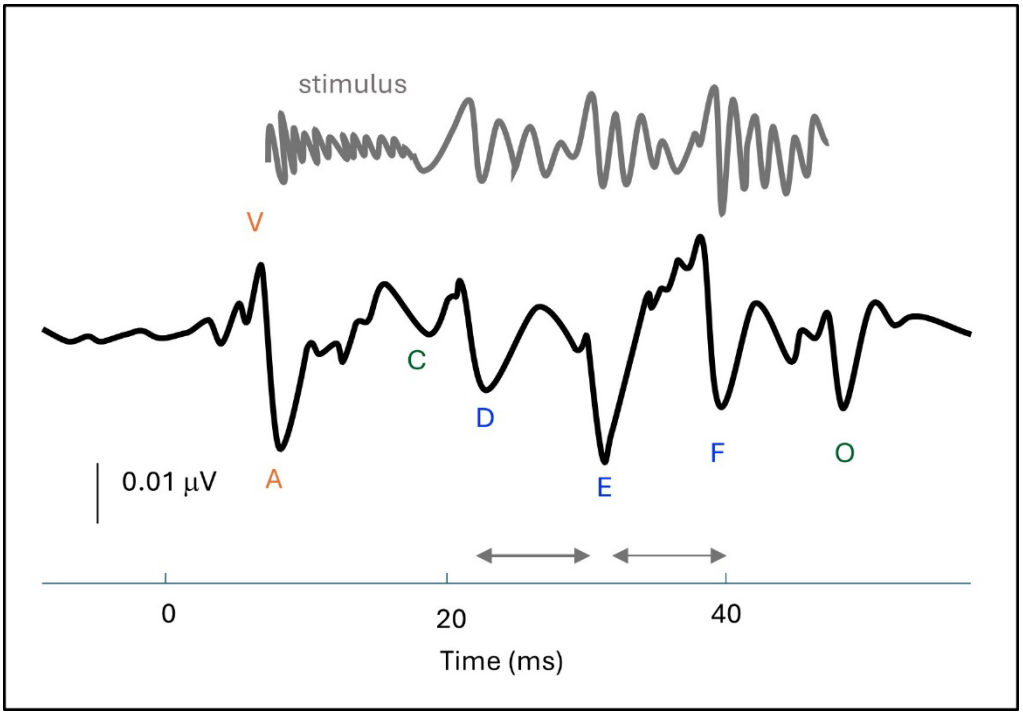

Figure 2.3 Time domain representation of the cABR in response to /da/. *This figure illustrates cABR to the stimulus /da/. The stimulus waveform is shown in grey, while the response waveform is depicted in black. This representation is adapted from the work of Skoe and Kraus (2010).*

The speech-evoked ABR using the CV syllable has been investigated in several studies but has yet to be used in clinical settings. Novis and Bell (2019) reported that the cABR to the /da/ sound exhibited <100% detection across normal-hearing participants (N = 16). Moreover, the morphology of the /da/ response was highly variable; some waves showed fair detection rates, while others were poor. In contrast, responses to clicks in the same group of subjects showed a 100% detection rate for wave V. The feasibility of using speech ABR through HAs was investigated by BinKhamis et al. (2019), who measured the cABR to the /da/ sound in HA users (N = 98). Their results showed significant earlier latency of all peaks and larger amplitude of peaks V, A, D and F in aided responses. Although the responses were sensitive to aided measures, they were still unsuitable for clinical use due to high variability between subjects in detectable peaks. The onset peaks V and A and the offset peak O were missing from a high number of participants (BinKhamis et al., 2019). Further, the /da/ sound lacks natural continuous speech's frequency range (bandwidth).

2.3.3 Speech CAEPs

The behavioural speech tests used to evaluate HA fittings are unsuitable for infants under six months, emphasising the need to find an objective measure (Dillon, 2005). The NAL group developed one objective speech detection outcome measure of HA performance in infants

(Purdy and Kelly, 2001). This test uses stimuli with speech features that can assess speech detection and accordingly provide a realistic measure of hearing performance when using HA (Punch et al., 2016). Further, the test is not limited to the peripheral system but assesses the whole auditory pathway up to the cortex.

The current CAEP protocol developed by the NAL group is used by Australian Hearing—a government organisation that provides audiological services for paediatrics using HAs or CIs. The stimuli used are the short consonant sounds /m/, /g/ and /t/, which represent low-, mid- and high-frequency speech sounds, respectively (Dillon, 2005). Stimuli are presented in the free-field via loudspeaker using two stimulation levels (Punch et al., 2016). The initial presentation level is 65 dB SPL, followed by either 55 dB or 75 dB SPL. Responses are quantified within an average of 100 artefact-free epochs and presented in the time domain, representing the primary ALR responses, namely, P1, N1 and P2 (Dillon, 2005). The morphology of adult responses exhibits all three components (P1, N1 and P2), while only a single dominant peak can be detected in infants. The test is performed using a HearLab device, which was introduced as part of Australian Hearing's clinical practice in 2011. A three-electrode configuration is used: active at Cz, referenced to the left mastoid, and ground at the forehead (Carter et al., 2010), This configuration is used to record the responses. The first aided CAEP evaluation is typically performed during the second follow-up appointment, approximately eight weeks after the initial fitting (Punch et al., 2016).

CAEP responses assess the aided detection of speech sounds; however, it is important to consider how the responses reflect the actual benefits of the HA. The correlation between the CAEP response in infants and the score of a parent questionnaire was examined by Golding et al. (2007). This study included 28 infants and children who were frequent HA users, based on their parents' reports. The PEACH scoring system was used as a subjective measure of HA benefit. A significant positive correlation was found between the detectable CAEP responses and age-corrected PEACH scores. To examine the advantage of using CAEP further, the researchers tested the correlation between the ABR/ECochG threshold and age-corrected PEACH scores and found no correlation. This finding shows that speech CAEPs are better than standard ABR in predicting HA benefits. However, the correlation between CAEP and PEACH scores is relatively low ($r = 0.47$). Additionally, the variation in age-corrected PEACH scores was only partially explained by the models, suggesting that additional variables affect how well PEACH scores are predicted (Golding et al., 2007).

The CAEP protocol was reviewed by Punch et al. (2016) to investigate the test's feasibility in actual clinical settings, including the clinician's experience, perceptions of test protocols and how those perceptions influence management decisions. The data consisted of files from 87

infant cases in which CAEP had been recommended. A survey of 32 paediatric audiologists was also conducted. The results showed that approximately 80 % of infants were tested, and any decisions not to test were related to reasons outside the clinician's control (e.g., unattended appointments, unavailable equipment or other infant conditions such as otitis media). The test protocol was deemed clinically feasible in terms of time, as it could be implemented within the 60 min to 90 min of a standard clinical audiology appointment with few barriers faced by the clinician. The test results influenced the management plan, and most parents were satisfied with the test technique—especially when behavioural tests were inapplicable.

The CAEP protocol shows potential as an outcome measure to verify how HAs process speech sounds. However, there are some drawbacks. The absence of CAEP in infants is not always linked to decreased audibility; clinicians must be cautious when interpreting CAEP results (Gardner-Berry et al., 2016). Another drawback relates to how HAs process speech stimuli used in the CAEP test protocol. The speech stimuli are presented with interstimulus intervals (ISIs) ranging from 1 s to 2 s, limiting their resemblance to natural running speech (Easwar et al., 2012). HAs are programmed to rapidly and continuously process speech input, which can result in different output levels—both overall and at the onset—for phonemes presented with ISIs or time intervals of 1 s to 2 s, compared to the same phonemes in natural speech. Easwar et al. (2012) investigated how non-linear HAs process phonemes presented in isolation with 1 s to 2 s intervals in between compared to phonemes in natural running speech. They reported similar HA outputs with the two phonemes conditions among roughly 75% of subjects, while the remaining subjects exhibited lower outputs in isolated phoneme conditions. The significantly lower phoneme output level in 25% of the cases indicates that the aided responses to phonemes in isolation might underestimate actual speech audibility.

In a recent study by Visram et al. (2023), the sensitivity, repeatability and feasibility of aided CAEP in infants to repeated speech stimuli were investigated. A notable strength of the study is its large sample size of 103 infants, providing a valuable reference for future research. The infants, aged 3 to 7 months at the first visit, underwent EEG recordings using a single channel between FPz and the right mastoid. Two synthetic speech stimuli, mid-frequency and mid-high frequency, were presented at dB sensation levels (SL) ranging from 0 dB to 20 dB SL. Testing was repeated after seven days. The findings demonstrated strong clinical feasibility, with a 99% completion rate and an average testing time of 15 minutes. For SL >10 dB, the detection rates were 94% for mid-frequency stimuli and 79% for higher frequencies after two tests. However, in a single test, these rates dropped to 80% for mid-frequency and 60% for high-frequency stimuli, raising questions about the clinical feasibility of relying on a single test. One limitation of the study is the use of synthetic speech stimuli, which may not fully represent real-world speech

environments. However, the high completion rate with a testing time of 15 minutes provides a practical estimate for implementing similar tests for infants in clinical settings.

2.3.4 CE-Chirp

A recent study proposed the use of speech-like stimuli to evoke an ASSR (Laugesen et al., 2018). The ASSRs evoked by a narrow-band CE-Chirp showed a significantly larger response amplitude than the conventional amplitude and frequency-modulated tone. However, the HA might treat the stimuli as noise when using modulated tones or noise to elicit the aided ASSR responses. Therefore Laugesen et al. (2018). proposed a new modification for the narrow-band CE Chirp that might better estimate actual HA processing. The study used modified four narrow-band CE-Chirps with centre frequencies of 500 Hz, 1 kHz, 2 kHz and 4 kHz to resemble the envelope of the ISTS. This was achieved by first filtering the ISTS stimulus through four bandpass filters, each with a central frequency corresponding to the NB CE chirp. The envelope of each ISTS band was then extracted and applied to the corresponding chirp. Comparisons of the standard narrow-band CE-Chirp and the ISTS-modified version revealed that the latter resulted in lower ASSR amplitudes with a similar detection rate, as well as an increase in detection time of about 1 min (Laugesen et al., 2018). As the most critical clinical factor is detection time, and given that speech-modified CE-Chirp has only an additional minute in recording time, it is possible to use it to assess and measure aided outcomes. ASSRs to the modified CE-chirp were detectable in normal-hearing adults, but it is yet unclear whether this approach can produce detectable responses in adults with hearing loss or in infants. While these stimuli resemble the speech envelope, they lack the temporal fine structure of natural speech. In addition, the question of whether the ASSR measured here can be correlated with behavioural speech tests remains unanswered.

2.4 Auditory evoked responses to running speech

The AEP is very low in amplitude compared to ongoing EEG activity. The averaging method has been successfully used to improve the visualisation of the evoked responses. However, averaging requires discrete and short stimuli to elicit responses, which does not reflect the kind of sounds encountered in a real-life environment (Lalor et al., 2009). Several new approaches to analysing EEG responses have emerged with recent developments in modern computation capabilities. These approaches overcome the need for averaging and provide the possibility of using continuous speech as a stimulus. However, the neural processes reflected in responses to natural speech are still being explored. It is inconclusive whether the measured responses

reflect just the acoustic features of the stimulus or if they also assess speech intelligibility (Zou et al., 2019).

A considerable portion of the literature analyses such stimuli using two methods. One such method estimates the correlation function between neural responses and specific features of speech, such as the speech envelope (Aiken and Picton, 2008b, Kong et al., 2014, Forte et al., 2017, Petersen et al., 2017). Another method predicts the auditory system transfer function response using ongoing neural responses (Lalor et al., 2009, Lalor and Foxe, 2010, Ding and Simon, 2012b, O'Sullivan et al., 2015). The next section reviews the recent literature on analysing EEG responses to continuous running speech.

2.4.1 Cross-correlation

Cross-correlation is a signal processing method of objectively determining how two series correlate at different time lags (Petersen et al., 2017). The peak indicates how strongly the signals are correlated when optimally adjusted for any time-shift between the signals. When analysing EEG responses to speech stimuli, specific features (e.g., the speech envelope) are correlated with the ongoing EEG to calculate the cross-correlation function. The statistical significance of the cross-correlation function can be used to determine the presence of a response, and the maximum time lag indicates its latency.

Aiken and Picton (2008b) were among the earliest researchers to attempt to implement cross-correlation for measuring cortical responses to continuous speech. For normal-hearing adults (N = 10), the cross-correlation function was estimated between averaged EEG responses to 100 repetitions of six different sentence roots and the envelopes of the sentences. The sentences were presented in five blocks, each lasting approximately 12 min. Each sentence length ranges between 2.64 s – 3.64 s. A windowed cross-correlation method was used to compute the correlation between the stimulus envelope and cortical response at every sample point. Each 500 ms time window was correlated with 500 ms of the response, with delays between 0 ms and 300 ms in 4 ms intervals. The correlation between the response waveform and the transient response model was computed as an alternative model to compare. The transient response model resulted from the convolution between the P1-N1-P2 complex response to sentence onset and the first derivative of the stimulus envelope. The envelope derivative includes more acoustical landmarks as it shows the rate of amplitude change in the envelope (Chalas et al., 2022). Using the negative derivative in the convolution caused an inversion of the P1-N1-P2 complex. To prevent this, the absolute value of the most negative deflection in the derivative was added to the entire calculation (Aiken and Picton, 2008b).

The results showed a significantly higher mean correlation between the averaged responses waveform and the stimulus envelopes than was obtained from the transient response model. This finding indicates the possibility of using cross-correlation to detect cortical responses to sentences. Although the method successfully detected significant cortical responses in all participants, it might not be ready for clinical settings (Aiken and Picton, 2008b). Due to the time needed for electrode placement and removal, recordings with many channels are impractical for clinical use (Aiken and Picton, 2008b). The study used repeated short sentences, which might not resemble continuous running speech, but the detection method can still be used as a foundation for future work. One approach to investigating the clinical feasibility of the cross-correlation method could involve using only a single EEG channel.

The cross-correlation method was further tested by Kong et al. (2014) to track the effect of attention on cortical responses to the speech envelope in quiet and competing speaker conditions. Cortical responses to a continuous stream of speech were recorded in normal-hearing adults (N = 8). The speech stream was divided into segments, each with a duration of 50 s. In the quiet condition, participants were asked to either actively attend to the speech or passively listen while watching a movie. Participants were asked to attend to either a male or female speaker in the competing speaker condition. The cross-correlation was calculated between their EEG response and the entire speech segment (50 s) at time lags ranging from 200 ms to 600 ms at 4 ms intervals. The chance-level confidence interval was estimated from 1000 bootstrap resamples using randomly selected segments of stimulus and response.

In the study, they found that the peaks of the cross-correlation function roughly corresponded to the peaks of the CAEP, specifically N1, P2, and N2. Consequently, they used these peaks for their analysis. In the quiet condition, no significant difference in the correlation value, nor the latency of peaks in the correlation function, were observed between the active and passive listening conditions. In the competing speaker condition, the attention effect showed a significantly higher N1 peak amplitude of the cross-correlation function. These findings suggest that direct cross-correlation may be used to indicate the attentional state of a participant using continuous speech as a stimulus. The cortical responses were also detectable even when the participants were not attending to stimuli. It should be noted that the paper only presents group averages without showing the ability to detect responses in individual subjects. To develop this method into a clinical tool for diagnosis or HA verification, it must be capable of providing robust results for each individual, not just on average.

Forte et al. (2017) used the cross-correlation approach to detect brain responses to continuous speech at the subcortical activity level. Specifically, brainstem responses were quantified by measuring correlations between EEG responses and the fundamental waveform of continuous

speech. Computing the fundamental waveform of the speech involved several steps, beginning with low-pass filtering the signal at 1500 Hz, detecting the fundamental frequency of the voiced parts, and then using the Hilbert-Huang Transform to extract the fundamental waveform. The method also included a control condition where brainstem responses were tested with inserts placed far from the ear. This condition represented the no-stimulus condition, and no significant correlation was indicated. In the quiet condition with no noise presented with the stimulus, a correlation between speech fundamental frequency and brainstem response was observed at a latency of 9 ms in 87.5% of normal-hearing participants (N = 16).

The method was further investigated in the same study to measure modulation by attention in the presence of a competing speaker. The responses were more substantial in all subjects (N = 16) when the participants attended to the target speaker, with statistically significant increases in eight of them. These results demonstrate the potential for using a correlation-based method to show the presence of speech features at the brainstem level. The study was also the first to detect the effect of attention at the brainstem level, contradicting the common assertion that only cortical responses are enhanced with attention. Since the suggested approach does not call for short repetitive stimuli and prevents any brain adaptation, it may thus open up new avenues for research into the role of the brainstem in objectively assessing the perception of continuous speech with and without an HA. However, assessing AEPs at the cortical level might provide better insight into the neural processing of speech stimuli compared to the brainstem level (Lightfoot and Kennedy, 2006).

2.4.2 Temporal response function

Other studies have applied an alternative approach to analysing continuous speech (Lalor et al., 2009, Lalor and Foxe, 2010, Crosse et al., 2016). This new approach was built on the transient responses model proposed by (Aiken and Picton, 2008b). The new approach models the auditory system function and uses it to either predict EEG responses (forward modelling) or reconstruct the speech stimulus (backward modelling). The forward model assumes that the output of the auditory system, y(n) (EEG responses), is the convolution of some feature of the input stimulus, e.g., the envelope, *x(n)*, combined with unknown impulse responses, *w(n)*, and noise (Lalor et al., 2009).

Equation 2.1

$$y(n) = x(n) * w(n) + noise$$

where * indicates convolution. Given the known stimulus *x(n)* and the measured EEG responses *y(n)*, the impulse responses *w(n)* (usually referred to as the temporal response function – TRF in this field) are estimated via least-square estimations performed on a sliding window of the

stimuli and the recorded responses. This process can be analytically performed using the following equation:

Equation 2.2

$$w = \langle x_t x_t^T \rangle^{-1} \langle x_t y_t \rangle$$

Here $\langle x_t y_t \rangle$ represent the cross-correlation matrix between x_t (stimulus values within a specific time window around t) and y_t (the EEG responses at *t*), whilst *T* refers to the transposition of the matrix. The autocorrelation matrix of the stimulus $\langle x_t x_t^T \rangle$ is inverted. An essential step before autocorrelation matrix inversion is regularisation (Lalor et al., 2009, Crosse et al., 2016). Regularisation most commonly involves adding a scaled identity matrix to the autocorrelation matrix to ensure it can be inverted (Crosse et al., 2016). This step involves incorporating additional information to address ill-posed estimation problems and moderate the risk of overfitting. Ill-posed estimation problems arise when inverting the autocovariance matrix $\langle x_t x_t^T \rangle$, which can be numerically unstable, causing high variance in the estimate. This instability is more likely with non-white stimuli like speech, as it may be singular. To address this, a bias term is added to reduce variance and minimize overall estimation error. By doing this, the estimate may become more biased, but it will reduce the variance and result in a lower overall estimation error (Lalor et al., 2009).

This method was initially named the auditory evoked spread spectrum analysis (AESPA) (Lalor et al., 2009). However, TRF has recently become the term more commonly used to describe the final model of stimulus-response mapping (Crosse et al., 2016). Specifically, the model explains the relationship between some features in the stimulus and the neural response. Shannon et al. (1995) achieved nearly perfect speech recognition scores with broadband noise stimuli modulated by the speech envelope, confirming that the envelope in a small number of frequency bands holds sufficient information to permit speech recognition. Therefore, several studies since have used the speech envelope to estimate the transfer function (or equivalently the impulse responses function) of the auditory system (Lalor and Foxe, 2010, Ding and Simon, 2012b, O'Sullivan et al., 2015, Vanthornhout et al., 2019, Decruy et al., 2020). Other studies have used speech stimulus representations such as spectrograms, phonemes and phonemic features (Di Liberto et al., 2015, Drennan and Lalor, 2019, Gillis et al., 2023) as the input in estimating TRFs.

Originally, the TRF was based on forward modelling, which compares predicted responses to actual EEG responses (Crosse et al., 2016). This model is assessed by computing the correlation coefficient between the actual and predicted EEG responses (see Figure 2.4). The analysis is univariate because data from the other EEG channels are discounted, and the model processes each channel separately (Crosse et al., 2016). Forward modelling can provide

information about the spatiotemporal characteristics of neural responses (Etard et al., 2019). In an early study, Lalor et al. (2009) observed a significant correlation between averaged repeated responses and estimated impulse responses using forward modelling. With a high SNR in the speech stimulus, the new method performed better than standard averaged responses, when the stimuli were continuous. Signal averaging was performed for the continuous stimulus by averaging the time-locked responses to positive intensity increases in the stimulus when it exceeded 0.5 μV. The results indicate the new method's flexibility when dealing with continuous changes in the acoustic characteristics of the stimuli. A more recent study found that the TRF waveform produced by the forward model had a similar morphology to the averaged late responses following repeated short stimuli (Vanthornhout et al., 2019). Both studies show that the model can represent the responses in a way that can be interpreted visually through the morphology, latency and amplitude of the waveform.

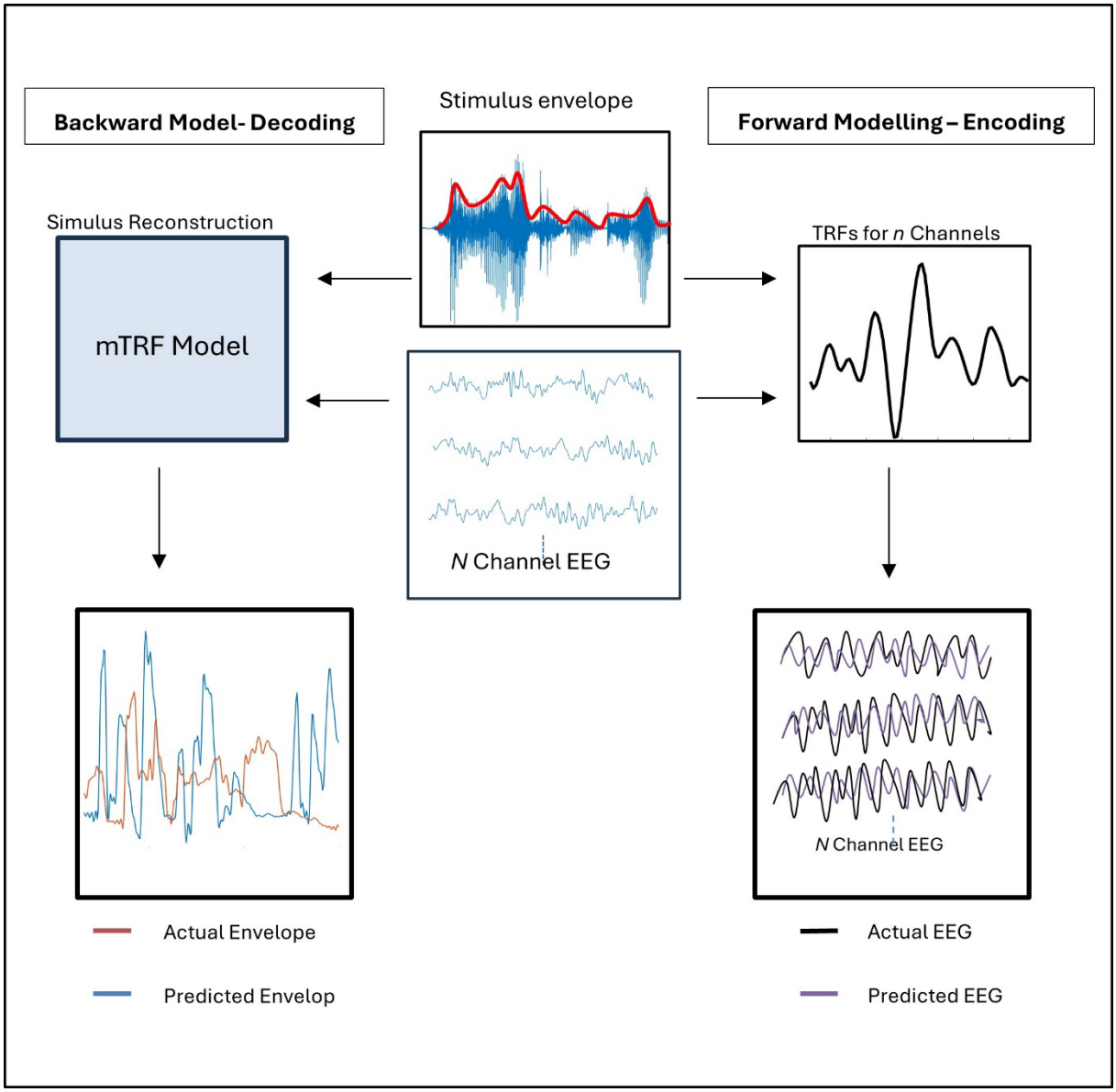

Figure 2.4 Diagram of the forward and backward modelling approaches used in the mTRF toolbox*. The figure presents the processing steps of the forward TRF (encoding) and the backward TRF (decoding). In forward modelling, the envelope stimulus (red line) is processed through the estimated TRFs from each channel, resulting in a predicted EEG waveform. The predicted EEG is then correlated with the actual EEG, yielding a correlation value for each EEG channel. In backward modelling, the EEG data from all recording channels are processed through the mTRF to reconstruct the stimulus envelope, which is then correlated with the actual envelope, resulting in a single correlation value. Adapted from Crosse et al. (2016).*

Backward modelling implements a reversed direction in terms of response–stimulus mapping (Crosse et al., 2016). It reconstructs or decodes specific speech stimuli features, most commonly the speech envelope, from recorded neural responses (see Figure 2.4). It can be made as an inherently multivariate approach (multiple EEG channels to improve the prediction of the speech envelope) using multiple electrodes, so selecting channels to be included in the analysis is not required (Mesgarani et al., 2009). Instead of preselecting channels, the model optimally weights each channel through the TRFs to provide the best approximation of the

predicted envelope (blue in Figure 2.4) to the true envelope (red in Figure 2.4), providing weights for each electrode based on the SNR (Pasley et al., 2012, Etard et al., 2019). The correlation coefficient values resulting from this backward model are usually between (r = 0.05 - 0.25) (Vanthornhout et al., 2018, Gillis et al., 2022a), which is higher than the predictions of the (univariate) forward model (r = 0.03 - 0.10) (Di Liberto and Lalor, 2017, Prinsloo and Lalor, 2022). The correlation value is currently the main parameter used to detect the presence of responses; to the best of the author's knowledge, the morphology of the TRF has not yet been used to detect the significance of the response.

Moving forward into how these methods can be implemented clinically, the next two sections will discuss the clinical feasibility and application of detecting neural responses to continuous speech stimuli.

2.4.3 Consideration of the clinical feasibility of using AEP to continuous speech

Several factors must be considered when assessing the clinical practicality and feasibility of specific tools. The first factor is the availability of testing equipment. The second is the time required for setup and for recording reliable responses. Additionally, attention and language are crucial when measuring cortical responses to speech in infants. This section reviews these considerations, focusing on the XCOR and TRF approaches to analysing cortical responses to continuous speech.

In laboratory settings, recording neural responses at a cortical level requires multichannel systems with 30 channels or more (Hall, 2015). The higher the number of channels, the more preparation and cleaning are required; equipment is also more expensive (McArdle and Hnath-Chisolm, 2015). The pathway by which a clinical protocol for recording cortical responses usually progresses is through first using a high number of channels, then moving towards lower numbers to reach clinical feasibility (McArdle and Hnath-Chisolm, 2015). All previous studies of cortical entrainment to speech envelopes have used multichannel recording systems to document EEG responses. The lowest number of channels used to record an EEG response was a 32-channel system with adult participants (Vanheusden et al., 2020), and 27 channels with infants (Jessen et al., 2019). Even though most studies use a high number of channels, lower channel numbers are required to measure significant responses. For example, Di Liberto et al. (2015) reported a significant correlation for the forward model by averaging correlations across 12 channels; no significant differences in correlation values were observed between the channels. Etard et al. (2019) demonstrated that EEG data from the CPz channel, in response to speech fundamental frequencies, showed significant attention effects on amplitude modulation at the subcortical level. This finding indicates that using a single EEG channel was

sufficient to detect the attention effect on the cortical response. None of these studies only used a single channel to investigate at the cortical level. This leaves the question of whether one channel can record and detect cortical responses to running speech yet to be addressed.

Recording time is another crucial factor in designing a test for clinical settings. Only a few studies in the literature have studied the influence of recording time when measuring evoked responses to speech. In the study by Di Liberto et al. (2015), roughly 72 min of a recording was needed to decode phonemic-level processing at the cortical level. Using the same method of forward modelling, researchers also investigated the minimum time required to measure significant responses (Di Liberto and Lalor, 2017). For each subject, the response was predicted using a subjective-specific model or a generic model. The subjective-specific model was a multivariate temporal response function (mTRF) fitted to data from the tested subject. The generic model was the average of subjective-specific mTRFs from all other subjects. The authors found that at least 30 min were required to measure responses using the subject-specific model. However, just 10 min of data from each subject was sufficient in the generic model to show a significant correlation. Jessen et al. (2019) studied both subject-specific and generic models, finding that 5 min of an EEG recording was adequate to show a significant correlation with the forward model in infants. However, the experiment was based on multisensory stimuli (audio and visual) that related to each other (a cartoon film). The audio speech envelope was extracted and used to estimate the audio TRF, but it was unclear whether the audio responses were influenced by the visual information presented. To ensure the reliability of the measured speech responses, researchers should use speech stimuli without related visual inputs.

Neural processes reflected in responses to natural speech are still being explored. It has yet to be confirmed, for example, whether they reflect acoustic onsets, comprehension or intelligibility. One investigation approach is to correlate neural responses to a behavioural speech test. Speech recognition tests such as BKB have been used in several studies to verify speech understanding in the presence of background noise (Vanthornhout et al., 2018, Lesenfants et al., 2019, Vanheusden et al., 2020). Children cannot be tested for behavioural speech perception before they develop the necessary motor skills to respond (e.g., picture pointing) or language skills to comprehend (Eisenberg et al., 2005). Infants may not yet possess receptive language, such as following simple directions or responding to questions. Therefore, speech perception tests should focus on detecting the presence of speech perception rather than predicting intelligibility.

Tracking neural entrainment to native and unknown languages presents a way to test if audibility or further aspects of language processing are being assessed. Etard and Reichenbach (2019)

recorded EEG responses to continuous running English and Dutch speech stimuli presented to native English speakers who did not understand Dutch (N = 20). The speech was present in quiet conditions and with three different levels of background noise. The researchers used backward modelling to track neural cortical responses to the speech envelope; no significant difference in correlation value was found in the quiet condition between the two languages. The speech amplitude modulated spectrum was found to be highly similar between English and Dutch (Varnet et al., 2017), explaining the significant responses to Dutch. In background noise, neural tracking of English speech was significantly higher than in Dutch. The results suggest that while significant neural tracking of foreign languages can occur with high SNR, detection may take longer than when working in the subject's native language.

Another possibility would be to use universal speech stimuli like the ISTS (Holube et al., 2010). The ISTS is a recording that remixes natural speech from six languages—English, Arabic, Mandarin, French, German, and Spanish—to produce a non-intelligible sound with several characteristics of natural speech, such as modulation frequency and fundamental frequency. The aim of developing such a signal was to provide an international stimulus that could be used to measure HA gain in real life. Using such stimuli to track cortical responses has not yet been investigated. Such stimuli can provide a universal method of tracking neural responses to continuous running speech as it is language-independent.

Auditory late responses are affected by the subject's state of attention (Hall, 2015). The N1 and P2 peaks amplitude were found to be more prominent when the subjects attended the stimuli (Picton and Hillyard, 1974). Controlling subject attention to the stimuli is essential when measuring cortical responses to speech. However, this presents a challenge when testing infants. Vanthornhout et al. (2019) investigated the neural tracking of the speech envelope when the subject actively attended to the stimulus or passively listened while watching a silent movie. They estimated the SRT under both conditions based on the correlation between the actual and reconstructed speech envelopes. Two lists of 20 sentences were presented to normal-hearing adults (N = 19) in conditions of quiet and several SNR levels. A significantly higher correlation was observed with attention in the low SNR conditions, whereas no difference in correlation between active and passive conditions was apparent when the SNR was high. Kong et al. (2014) also found that in the quiet condition, cortical responses to running speech were detectable with passive listening. Both studies demonstrate the feasibility of using the passive condition when tracking neural responses to the speech envelope. Watching a movie during testing, for example, can be more engaging for participants, especially when the subjects are young children or infants.

2.4.4 Applications of AEP to continuous speech

Tracking neural responses to continuous stimuli has become increasingly popular in recent years, with several applications for diagnosing impairment of the auditory system and attentional selection. Presenting continuous speech as a stimulus is more engaging and realistic for participants than brief tone stimuli (Power et al., 2012). Several researchers have used the TRF to investigate brain-selective attention to specific speech streams (Power et al., 2012, O'Sullivan et al., 2015, Vanthornhout et al., 2019). Further, many studies have used the model to measure SIN performance objectively (Ding and Simon, 2013, Vanthornhout et al., 2018, Lesenfants et al., 2019) and assess auditory and linguistic processing (Gillis et al., 2022b). Other applications, namely aided response (Petersen et al., 2017, Decruy et al., 2020, Gillis et al., 2022a) and infant speech detection (Jessen et al., 2019, Attaheri et al., 2022a), have been the subject of a recent investigation. As this project aims to assess the feasibility of using neural tracking methods for continuous speech in infant-aided testing, some critical works in the area of aided responses and infant testing will be discussed below.

2.4.4.1 Aided speech detection

Previous studies have successfully observed aided neural tracking of speech envelopes in listeners with hearing loss (Petersen et al., 2017, Vanheusden et al., 2020, Alickovic et al., 2021). Petersen et al. (2017) measured the cortical responses using the cross-correlation approach. They investigated the effect of hearing loss on neural tracking in 27 elderly participants whose hearing levels (pure tone average) ranged from 11 dB to 73 dB HL. Also, the influence of attention was measured at different SNR levels. Responses were detected from both attended and ignored stimuli at the group level. Although significant differences were indicated between attention conditions, this difference was reduced with severe degrees of hearing loss. From this study, it is unclear whether significant neural tracking of speech envelope was seen in all subjects, which questions the clinical feasibility of such a method in assessing aided speech responses at the individual level.

Vanheusden et al. (2020) proposed measuring cortical entrainment to evaluate HA performance objectively. The study compared correlation values between actual and reconstructed speech envelopes under aided and unaided conditions. Participants (N = 17) with mild-to-moderate hearing loss were presented with a continuous running speech from an audiobook that lasted 25 min. The authors found no significant difference in reconstruction accuracy between aided and unaided responses, which was explained by the good audibility and high subjective score in the unaided condition. However, the results also confirmed that it was possible to measure cortical entrainment despite non-linear HA processing effects on the speech envelope. These

findings laid the foundation for using such a method to improve objective HA outcome measures. However, since the presentation level was relatively high (70 dBA equivalent sound pressure level (LeqA SPL)) and participants had mild to moderate hearing loss, the HA may have had minimal impact on the speech signal. In more realistic settings with lower sound intensities, HAs would provide greater amplification, resulting in more significant compression of the stimulus. Compression has been shown to affect the speech envelope (Stone and Moore, 2007, Chinnaraj et al., 2021), which could influence the detection of responses to continuous speech, as the envelope is a key feature used in the model.

More recently, Alickovic et al. (2021) investigated the effect of the HA noise reduction (NR) on the neural tracking of the speech envelope by measuring the reconstruction accuracy of the backward model. The sample included 34 subjects with bilateral mild to moderately severe hearing loss, who were experienced HA users. The study examined early AEPs, defined as neural responses occurring at latencies of less than 85 ms, and late AEPs, characterised by latencies between 85 ms and 500 ms, which are associated with higher-level auditory cortical processing. The objective was to investigate how the NR scheme affects the neural representation of a multi-talker auditory scene at distinct hierarchical stages of the auditory cortex. A significant effect of NR was observed in the early responses, leading to higher reconstruction accuracy of the model for the entire acoustic scene. For the late responses, the results showed an enhancement of the target speech and suppression of background noise. The findings of these studies demonstrate the potential of using aided EEG with continuous speech as stimuli, supporting the results of Vanheusden et al. (2020). Additionally, Alickovic et al. (2021) further validate the potential of tracking cortical to continuous speech for assessing the benefits of specific HA features, such as NR, in complex and noisy listening environments. Based on the studies discussed in this section, a possible direction for future research is to explore the clinical application of this approach, which will be discussed later in this chapter.

2.4.4.2 Infant testing

Most of the existing literature uses adults as participants. However, a study by Kalashnikova et al. (2018) measured the neural cortical tracking of temporal speech modulation in infants, making it the first to test such an age group using the TRF approach. One major obstacle to measuring speech encoding in infants is the difficulty in tracking their attention. The researchers examined the effect of using engaging infant speech and then compared this effect to the use of regular speech. Two speech types were presented to (N = 12) seven-month-old infants, namely, infant-directed speech (IDS) and adult-directed speech (ADS). Per the forward model, the results showed a significant correlation between EEG responses and their predictions in response to IDS, but not to ADS. The response was associated with a cluster of eight frontal

electrodes that exhibited significant correlation. The nature of the speech stimuli used in the study suggests that it is possible to attract infant attention and quantify cortical entrainment to continuous speech. However, the difference between the two stimuli might not be purely explained with attention as the two stimuli are acoustically different. Additionally, IDS tends to have longer pauses compared to ADS (Fernald et al., 1989). A more robust cortical onset response was found only after longer pauses in the continuous stimulus (Hamilton et al., 2018).

Infant testing was further investigated by Jessen et al. (2019) through multisensory (audio and visual) stimulation and a larger sample size (N = 52) of seven-month-old infants. An additional sample of (N = 33) adults was included in the study to compare against the infant test results. Significant correlations between the predicted and actual EEG responses using the forward model were measured in both groups using subject-specific and generic models. The subject-specific model was an mTRF fitted on data from only the subject tested, whereas the generic model included the averaged subject-specific mTRFs from all other subjects. The study results demonstrated the ability to track neural responses in infants with only 5 min of simultaneous sensory stimuli. However, it is unclear whether the correlation was significant for all subjects. The study also defined the characteristics of infant TRF morphology in comparison to adults as lacking the first positive peak (150 ms–250 ms) and having a longer positive peak (250 ms–500 ms). The TRF waveform is in line with the CAEP evoked by short speech stimuli in infants, which has one prominent positive peak between 100 ms and 300 ms (Ahissar et al., 2001, Dillon, 2005). This finding corresponds to the similarity between TRF response morphology and late AEP evoked by short stimuli, discussed in Section 2.4.2.

Since the start of the project, more studies have been investigating the neural tracking of speech envelopes in infants (Ortiz Barajas et al., 2021, Attaheri et al., 2022a). The study considers how infants develop the ability to process speech, focusing on the tracking of speech envelopes in the brain during the first months of life. Using a cross-correlation approach, the study compares neural tracking of familiar and unfamiliar languages. The study found that newborns show the ability of neural tracking for both types of language, which provides evidence that the ability to track speech envelopes does not reflect comprehension. Interestingly, the study found that in the 6-month-old group, neural tracking was only significantly detected for the unfamiliar language. The absence of neural tracking of familiar language at this age was also seen by (Kalashnikova et al., 2018) The explanation given in the study was related to the neural development of language acquisition, which shifts from focusing on syllabic units, crucial for speech perception from birth to the processing of phonemic units, which are key for learning words and grammar starting around 6 months. This shift moves from elements present in the speech envelope at 4 Hz–5 Hz to phonemic units around 30 Hz, reflecting a developmental progression in language acquisition. These findings highlight the dynamic nature of neural

adaptation in early language development and raise questions about the necessity of using intelligible speech to assess speech perception at this age.

The studies mentioned earlier involving infants demonstrate the feasibility of using the TRF method and cross-correlation to test infant cortical responses to real-life stimuli. Three important considerations from these studies can help design a new infant experiment. First, infant attention can be captured using IDS or visual stimulation. However, since it might be impossible to control attention in infants, the feasibility of measuring cortical responses to continuous speech with passive listening should be investigated. Second, EEG recording lengths can be as short as 5 minutes, though it is unclear if detection sensitivity is 100% and no data was given regarding the detection of the responses for individuals. Finally, the use of speech stimuli independent of the infant's language supports clinical feasibility.

2.5 Summary and Research Motivation

It has been established that an objective approach can overcome the limitations of subjective methods in difficult-to-test cases, such as infants. This literature review shows that AEPs were the first method of choice for objective hearing assessment. First, different forms of conventional AEP, using short stimuli like tones, were explained. Next, several signal-processing approaches for analysing EEG signals when using speech as a stimulus were discussed. Most of these approaches relied on averaging responses to repeated stimuli, which have limited environmental relevance and do not reflect the performance outside clinical settings. Therefore, the literature review subsequently explored the use of continuous speech.

This review concluded by exploring recent literature on measuring evoked responses to continuous running speech, elaborating on possible clinical applications and feasibility considerations. The literature reveals a clinical need for an objective outcome measure that resembles everyday listening (Grill-Spector et al., 2006, Summerfield et al., 2008). This aim can be achieved by analysing auditory evoked responses to speech stimuli. Aided cABR provides limited information about how HAs process natural speech, as they rely on repeated short speech stimuli to elicit responses. In contrast, HA input in everyday listening environments consists of continuous speech and various background noises, encompassing a broader range of acoustic features. While aided cABR can offer some insights into HA performance, it does not fully capture how HAs function in more realistic auditory settings.

The current clinical method for assessing speech-aided response objectivity is the NAL group’s CAEP system (Punch et al., 2016). The technique uses four short speech sounds to approximate the speech frequency range. Compared to cABR, the CAEP approach might provide more insight

into processing speech up to the cortex. However, using continuous speech would present a more realistic measurement of speech perception.

2.6 Gaps in Knowledge

The literature on measuring cortical responses to continuous running speech reveals gaps in current approaches that require further investigation:

- What is the detection rate of cortical responses to continuous speech using EEG data from just one channel? This would reduce the setup time, significantly benefiting clinical settings.
- What is the minimum time required to detect significant responses in individuals using the TRF method and cross-correlation?
- Is there a significant difference in detection sensitivity between XCOR and TRF?
- Can other TRF parameters, such as peak amplitude and latency, be used to detect responses? The literature indicates that correlation values are the only parameter currently used.
- Is there a significant difference in detection sensitivity and detection time between cortical responses to intelligible speech and language-independent speech-like stimuli?
- Is there a significant difference in the detection rate of cortical responses from a single channel between passive and active listening conditions?
- What is the amount of envelope distortion caused by the HA, and how does it impact the measurement of cortical responses to continuous aided speech?

2.7 Thesis Objectives

The key objectives for this project's contributions towards the overarching aims are:

1. Assess the viability of detecting cortical responses to continuous speech using EEG data from a single channel, as typically available in current clinical systems, by measuring detection sensitivity and detection time.
2. Investigate the detection sensitivity of different parameters by comparing XCOR and TRF, including correlation values and other morphological features.
3. Determine whether the intelligibility of the stimulus influences cortical responses to continuous speech.
4. Compare the effects of active and passive listening on detecting cortical responses from a single channel in order to determine the role of attention.

5. Measure changes in the stimulus envelope after processing by standard HAs (including compression), assess the extent of envelope distortion and evaluate its impact on cortical responses to amplified stimuli.
6. Compare cortical responses to continuous speech with and without HA processing to assess the feasibility of detecting aided cortical responses.

Chapter 3 Measuring Cortical Responses to Continuous Running Speech Using EEG Data From One Channel

This chapter presents the analysis of pre-existing data from a study by Vanheusden et al. (2020). The analysis tests the detection sensitivity of several parameters using cross-correlation and forward modelling methods. The EEG data from a single channel are used to detect the presence of responses. The results provide evidence regarding the best approach to use for further steps in this project. This chapter is now published in the International Journal of Audiology (Aljarboa et al., 2023).

3.1 Introduction

As highlighted in Chapter 2, using natural running speech can provide a more environmentally valid measure of speech processing, with the potential for hearing aid evaluation. It may also overcome the issue of neural activity suppression caused by stimulus repetition (Grill-Spector et al., 2006). Response to continuous speech can better evaluate how HA users perform in realistic speech-in-noise situations (Decruy et al., 2020). Measuring auditory evoked responses to continuous running speech can be accomplished in two ways. The first is to measure the maximum cross-correlation between the EEG response and specific features of the speech signal (BinKhamis et al., 2019). Previous work has shown that both the temporal envelope (Kong et al., 2014) and the fundamental frequency of speech (Forte et al., 2017) correlate with the EEG signal. The second approach is typically referred to as a TRF, which estimates the relationship (transfer function) between the features of input speech and the corresponding EEG (Lalor et al., 2009). Through the TRF approach, EEG responses can be estimated from the speech stimuli (forward modelling), or the speech envelope can be reconstructed from EEG responses (backward modelling). In most studies, the significance of the correlation between the actual response and the predicted or reconstructed response is used to detect the auditory processing of the stimulus.

When using forward modelling, Vanthornhout et al. (2019) employed detection based on the peak of the TRF filter to measure the effect of attention, finding that the resulting impulse responses were similar to those of CAEPs following short tone bursts. However, it is not well established whether features of the TRF filter waveform, such as peak amplitude or power, can be reliably used as parameters to measure the presence of responses. In general, both the

cross-correlation and TRF approaches can detect cortical responses to running speech and have been widely used (Di Liberto et al., 2015, Di Liberto and Lalor, 2017, Vanthornhout et al., 2019, Vanheusden et al., 2020). However, a comparison of detection sensitivity between the two approaches has not yet been made, nor has there been an exploration of how TRF filter waveform parameters, such as peak value or power, might compare for detection purposes.

Developing a clinical tool requires careful consideration of practical aspects, such as the duration of testing. Notably, two studies – Di Liberto and Lalor (2017) and Mesik and Wojtczak (2022) – measured the minimum time needed to detect cortical entrainment to continuous speech in normal-hearing adults. Di Liberto and Lalor (2017) examined the time required to obtain a significant correlation using the speech envelope and other phonemic features to estimate the forward TRF model. For each subject, responses were predicted using either a subject-specific model or a generic model. The subject-specific model was an mTRF fitted on data from only the subject tested, while the generic model averaged the subject-specific mTRFs from all other subjects. The findings indicated that 30 min or more of EEG data recordings were required to measure significant correlations using TRF-forward modelling with the subject-specific model – a duration impractical for clinical use. However, when a generic model based on recordings from nine previous subjects was used, only 10 min of recordings were needed. For this approach to be clinically feasible, such a generic model would need to be available in advance, and it remains uncertain how well a model based on normal-hearing subjects would generalise to hearing-impaired subjects.

More recently, Mesik and Wojtczak (2022) conducted a more intensive investigation into the effect of EEG data quantity on prediction accuracy, progressively increasing the data range from 3 min to 42 min across 11 different data quantities. A key finding of their study was that the performance of the subject-specific model plateaued after approximately 14 min of data when the speech envelope was used as the primary detection feature. Neither study, however, addressed how recording time varied between different participants, leaving uncertainty about the time required to detect responses at the individual level.

Another important requirement for clinical use is the availability and suitability of equipment. Detecting EEG responses to continuous speech is usually carried out using a multichannel recording system. In addition to increasing the setup and testing time, multichannel systems with 32 or more electrodes are costly and not typically available in audiology clinics (Ginsberg et al., 2023). For practical purposes, using a single channel for clinical measurement would be preferable. In the current study, the forward model was considered the most appropriate approach, as the model is estimated based on data from a single channel. However, the ability to detect single-channel responses to running speech has not been well explored in previous

works, nor has the position of the best single-channel been identified. An advantage of using multiple channels may be that the best-performing channel does not need to be selected in advance, removing the risk of only recording a sub-optimum single channel. As yet, a comparison of single-channel detection with multichannel has not been made for the purpose of speech-evoked responses.

For standard AEP measurements, cortical responses typically exhibit high amplitude in the frontal area. Aiken and Picton (2008b) suggest that the source of cortical responses to the onset of speech sentences is in the superior region of both temporal lobes. The Hearlab system protocol developed by Carter et al. (2010) to measure standard AEPs to short speech sounds uses the Cz location to detect cortical responses to speech sounds. Another common measurement location is Fz, as it has a good response amplitude but avoids the need to put electrodes above the hairline (Hall, 2015). A comparison of these single-channel measurement positions in terms of detection time for responses to running speech has not yet been made. An alternative could be to use multichannel measurement and select the best single channel for each subject. Although this would be more complicated for clinical use, it might be worthwhile if it achieves a significant reduction in test time. A possible downside to the approach is that selecting the strongest statistical parameter, such as correlation, from multiple channels will increase the values obtained both with and without responses present. This will, thus, raise the threshold beyond which any response can be deemed statistically significant. Hence, this will be addressed in the current work.

The objectives of the current study are: (1) to compare the sensitivity of different analysis methods and parameters in detecting cortical responses to continuous running speech using a single channel, and to determine whether cross-correlation or TRF parameters are most sensitive for detection; (2) to explore the minimum time required to detect significant cortical responses at the individual level; and (3) to investigate whether a single-channel approach is more effective than a multiple-channel approach, where the best-performing channel is used for analysis.

Previous studies have demonstrated that it is possible to measure cortical responses to continuous speech using multichannel EEG with cross-correlation (Kong et al., 2014, Petersen et al., 2017)and TRF approaches (Vanthornhout et al., 2019, Vanheusden et al., 2020). However, this study further investigates this area by focusing on detection using a single-channel approach and examining the individual recording time required to detect a significant response. The sensitivity of the analysis methods will be evaluated based on detection rate and mean detection time.

3.2 Materials and Methods

The EEG data in this paper was previously collected by Vanheusden et al. (2020). The EEG data were recorded from 17 native English speakers (11 males, 6 females, age 65 ± 5 years) with bilateral mild-to-moderate sensorineural hearing loss. Hearing levels were assessed using pure-tone audiometry. Table 3.1 shows the mean and the standard deviation (SD) of hearing levels at each frequency for both ears. All participants were regular HA users and were tested under aided and unaided conditions. They were awake and attentive to the stimulus during the sessions. For the current work, only the unaided condition was included in the analysis. The condition was chosen to exclude the possible complication of HA processing: the aim was to compare the performance of different detection parameters, not to further explore the effects of aiding. As such, using only the unaided data simplified comparisons between methods. If the identification of the best-performing method for unaided data is successful, the next step will be to test its sensitivity in detecting aided responses.

Table 3.1 Subject average hearing levels. *The table shows the pure tone average mean and standard deviation (SD) of the subjects.*

	Hearing levels (mean +/- SD)	
Frequency	Left ear	Right ear
200 Hz	23 dB ± 13 dB	29 dB ± 21 dB
500 Hz	23 dB ± 15 dB	27 dB ± 21 dB
1000 Hz	29 dB ± 16dB	32 dB ± 23 dB
2000 Hz	43 dB ± 19 dB	43 dB ± 21 dB
3000 Hz	54 dB ± 17 dB,	51 dB ± 20 dB
4000 Hz	61 dB ± 14 dB	61 dB ± 17 dB
6000 Hz	69 dB ± 19 dB	68 dB ± 24 dB
8000 Hz	68 dB ± 18 dB	65 dB ± 18 dB

This experimental stimulus was a 25-minute English narrative of continuous running speech from an audiobook. It was divided into eight segments of roughly 3 min played contiguously (Vanheusden et al., 2020). The speech stimulus was sampled at 44.1 kHz and low-pass filtered at 3 kHz. These parameters were chosen because of previous findings that frequencies above 3 kHz have no significant effect when measuring low-frequency cortical entrainment to speech (Vanheusden et al., 2020). The stimulus was presented at 70 dBA (LeqA SPL) through loudspeakers placed 1.2 m in front of the subject. Attention was ensured by asking questions after each segment. The EEG responses were collected using a 32-channel system (Biosemi, Netherlands, sampling rate of 2048 Hz). The electrodes were placed according to the 10–20 standard configuration and referenced to the averaged EEG responses over all electrodes.

3.2.1 Data analysis

The EEG responses were recorded from 32 channels, down-sampled to 128 Hz and referenced to the two mastoid electrodes. The EEG was filtered between 1 Hz and 30 Hz, covering the same band as the speech signal's envelope. In the first analysis, responses were detected using either the single Cz or the single Fz channel. In the second (multichannel) analysis, the channel with the highest correlation between the EEG signal and the speech envelope was selected from either all 32 channels or a subset of 6 channels. The six channels — Cz, Fz, F3, F4, FC1, and FC2 — were chosen based on their high correlation values and evidence from previous studies. This identified the frontocentral area of the scalp as having the highest correlation between the predicted and actual speech envelop (Di Liberto and Lalor, 2017). Recorded EEG data were analysed using two approaches: cross-correlation of the speech envelope with the EEG signal and the TRF forward model.

The TRF method was based on determining the impulse responses of the system by assuming a simple linear convolution between the input (the envelope of the speech signal) and the output (the EEG of each channel). The predicted EEG responses were computed from the speech envelope and the estimated forward model using the MATLAB TRF toolbox (Crosse et al., 2016). The speech envelope was calculated using the absolute value of the Hilbert transform. The envelope was down-sampled from 44.1 kHz to 128 Hz and filtered with a zero-phase filter (1 Hz – 30 Hz). Before TRF analysis, the EEG data and speech envelope were normalised by dividing by their SD. Artefact rejection was applied to the EEG data in blocks of 100 samples, with blocks that included values exceeding +/- 5 SD of amplitude being excluded from both the EEG data and the corresponding speech envelope.

Based on the results of the decoding, three main parameters were calculated to indicate response quality. First, the TRF peak value (TRF-peak) was measured in terms of the filter response peak-to-peak value. The second parameter was the power of the TRF (TRF-power): the average of the squared amplitude of the TRF filter across the window of interest (0 ms to 375 ms). The third parameter was the accuracy of the model's prediction (TRF-COR), given by the correlation coefficient (Pearson's r) between the actual and predicted EEG response in each channel. The TRF analysis code is provided in Appendix A.

A bootstrap statistical analysis method was used to determine whether the magnitude of a given parameter was significantly different from random variation. In evoked potential studies, bootstrapping has been used (Lv et al., 2007, Chesnaye et al., 2018, Vanheusden et al., 2019) to construct the null distribution for a specific statistical parameter of interest. The distribution is estimated by repeatedly and randomly drawing samples with replacements from the ongoing EEG data and calculating the statistical parameters for each sample (Chesnaye et al., 2018).

Here, this involved comparing the parameter values (e.g., the TRF-peak) obtained when the envelope and EEG were correctly aligned to the bootstrap distribution values when they are repeatedly and randomly misaligned (giving e.g., the TRF-peak* with * indicating bootstrap values). The bootstrap distribution was then tested to determine if it differed significantly from the original (with correctly aligned signals) value at a chosen significance level ($p < 0.05$). This bootstrap approach allows significance to be tested in each recording rather than testing for significant effects across the cohort.

For the TRF-peak and TRF-power parameters, 500 TRFs were calculated with random misalignment between the speech envelope and the EEG responses. The misalignment was always set to a minimum of 2000 samples (15.6 s) to prevent random alignments where the EEG coincided with the speech envelope. This produced a bootstrap estimate of the null distribution for each parameter (an estimate of what was expected when no responses were present will be denoted as TRF-peak* and TRF-power*, respectively). To find the significance of the TRF-peak, maximum and minimum values for each TRF-peak across a time window from 0 ms to 375 ms were determined. The minimum and maximum values were then sorted to determine the 2.5% thresholds (upper and lower), resulting in a 5% false positive rate. A response was considered present when the detected parameter value exceeded the 5% confidence interval (the TRF-peak values exceeded the upper or lower 2.5% of TRF-peak*). For the TRF-power, only the maximum 5% values were determined to detect the presence of a response. The response was considered significant if the TRF-power exceeded the upper 5% of TRF-power*.

For the TRF-COR parameter, the bootstrap null-distribution of correlation values was determined using again the 500 random misalignments of the EEG and speech envelope described above, to provide TRF-COR*. A response was identified as significant if the correlation value of TRF-COR exceeded the upper 5% of TRF-COR* (i.e., the probability of achieving this TRF-COR value under the null hypothesis was $p < 0.05$). The XCOR analysis code is provided in Appendix B.

The maximum of the cross-correlation function between the EEG and speech envelope (without using the TRF) was implemented for response detection. First, the EEG data and speech envelope were normalised such that they had an SD of 1. Artefact rejection was applied, as mentioned in the TRF analysis. The cross-correlation function between the aligned EEG data and speech envelope (XCOR) was calculated over a range of lags between the signals, from 0 ms to +375 ms. The maximum and minimum (the peak negative value) of the cross-correlation function were selected, and the significance of the peak or trough was determined again from the bootstrap distribution of XCOR*, following the same method as used for other parameters. Because the correlation between the EEG responses and the speech envelope was estimated,

both strong positive and negative correlations may be deemed significant (Aiken and Picton, 2008b). A response was thus considered present when XCOR exceeded the upper or lower 2.5% of XCOR*.

False positive rates were also tested to verify that the proposed methods of analysis were working as expected. Tests were applied to the four parameters (TRF-COR, TRF-peak, TRF-power and XCOR) using white noise as the input and output signals. Both analysis methods were tested for false positives using the time-reversed speech envelope as a control condition, as no significant correlation was expected between the time-reversed speech envelope and the EEG.

3.2.2 Single channel detection time analysis

The minimum time required for the response to be detected (detection time) was recorded for each subject. The purpose was to compare the performance of the four parameters outlined above (TRF-peak, TRF-power, TRF-COR and XCOR). The data were divided into segments of increasing length from 1 min up to 25 min in 1 min increments. For each segment, the detection parameter outlined above was determined to indicate whether a response was present. The detection time specified for each subject was the minimum recording period for which a significant response was first detected. In cases where a subject showed no detection, the detection time was 26 min (i.e., longer than the 25 min recording) to avoid missing data in the statistical analysis.

3.2.3 Analysis of the best channels from multichannel data

This study looked at detection for Cz and Fz (the most commonly used single channels in clinical practice). However, detection was also considered based on the selection of the 'best' channel (the channel with the highest correlation value for each subject) from multiple channels. Two sets of multichannel analyses were assessed: (1) across all 30 channels (32 channels were recorded, but the two mastoids were used as the reference); (2) across the frontal/central six channels that were most likely to contain a response (Aiken and Picton, 2008b). The results of these analyses were compared to the results from a single fixed channel. The detection rates were calculated at 5 min rather than in 1 min increments to reduce the analysis time. Two parameters were included in this analysis: TRF-COR and XCOR, as they showed the highest detection rates

The bootstrap significance test for each of the parameters was carried out in a similar manner to the one-channel analysis. However, to overcome the possibility of multiple comparisons producing false positive detections, the bootstrap analysis was then repeated for each of the

other channels. Rather than using the 5% confidence interval for a single channel and 500 random misalignments, the 5% confidence interval was calculated across 500 random misalignments, selecting the highest value from all channels of data to provide the null distribution for the significance of the result in the channel with the highest correlation to the stimulus envelope.

3.3 Results

The results of the false positive rate test showed approximately 5% in 1000 simulations. Table 3.2 presents the measured false positive rate for each parameter. All methods produced a false positive rate close to the expected value of 5% (or 50 out of 1000), which falls within the 95% confidence interval expected from the binomial distribution for a test at $p < 0.05$ with 1000 repetitions (i.e., between 37 and 64 false positives).

Table 3.2 Analysis of false positive (FP) rates using white noise. *The table represents the expected FP range for 1000 tests with white noise as the input and output signals for all testing parameters: TRF-Peak, TRF-Power, TRF-COR, and XCOR. Tests were calculated using the Binomial Distribution at $p < 0.05$.*

Parameter	TRF-peak	TRF-power	TRF-COR	XCOR
#Tests	1000	1000	1000	1000
FPs	49	41	46	53
% FP rate	0.049	0.041	0.046	0.053
Expected FP Range	37-64	37-63	37-64	37-64

3.3.1 Analysis of responses from single channels Fz and Cz

The first objective of this study was to explore the detection of cortical responses to continuous running speech using EEG data from one channel only. Figure 3.1 shows an example of the cross-correlation function results from one subject, and Figure 3.2 shows the TRF filter for the same subject. Figure 3.3 shows detection times for Fz (panel A) and Cz (panel B). For Fz, the best detection rates were achieved with the TRF-COR and the XCOR parameters, exhibiting detection in all subjects (17) after 17 min. With TRF-power and TRF-peak, 15 subjects exhibited detection after 19 min and 22 min, respectively. The Cz channel results showed detection in 16 subjects with TRF-COR and XCOR after 20 min and 21 min, respectively, and in 14 subjects for both TRF-power and TRF-peak after 18 min.

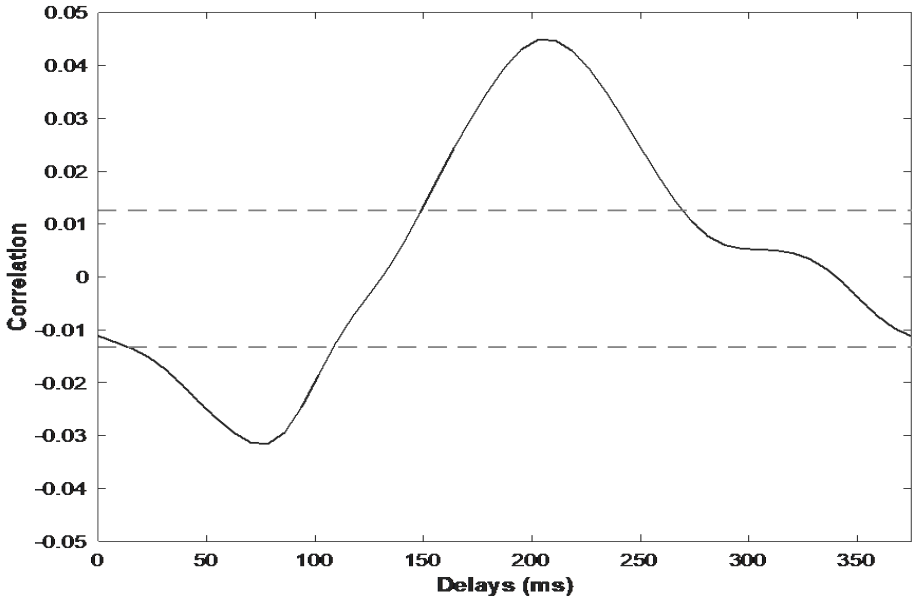

Figure 3.1 Example of the XCOR function for subject 1 (Channel Fz)*.The x-axis indicates the delay between speech envelope and EEG in ms. The y-axis indicates the correlation value. The area between the dotted lines indicates the 95% bootstrap confidence interval for no response data. The maximum value corresponds to the XCOR index for this recording and is statistically significant. Note that the time displayed includes the acoustic delay between the loudspeaker and the subjects, which, for a distance of 1.2 m, is approximately 3.5 ms.*

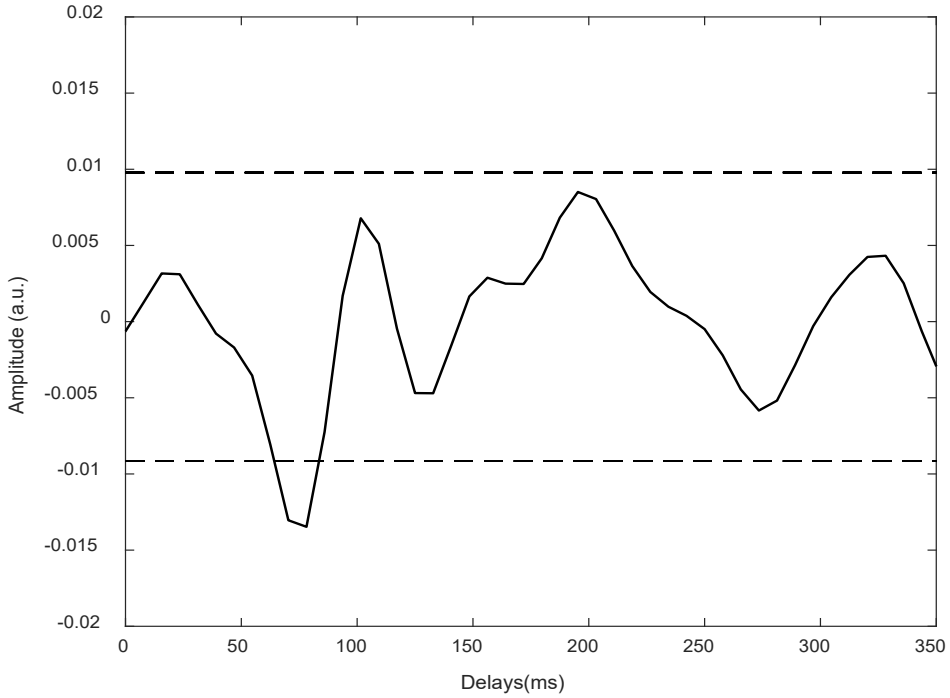

Figure 3.2 Example of the TRF-filter waveform for subject 1 (Channel Fz). *The x-axis indicates a delay between the envelope and EEG in ms. The y-axis represents the vertex potential, with upward deflections indicating positive polarity relative to the reference electrode. The magnitude of the TRF impulse response is shown on the y-axis. The area between the dotted lines represents the 95% bootstrap confidence interval for null-response data. Note that the time displayed includes the acoustic delay between the loudspeaker and the subjects, which, for a distance of 1.2 m, is approximately 3.5 ms.*

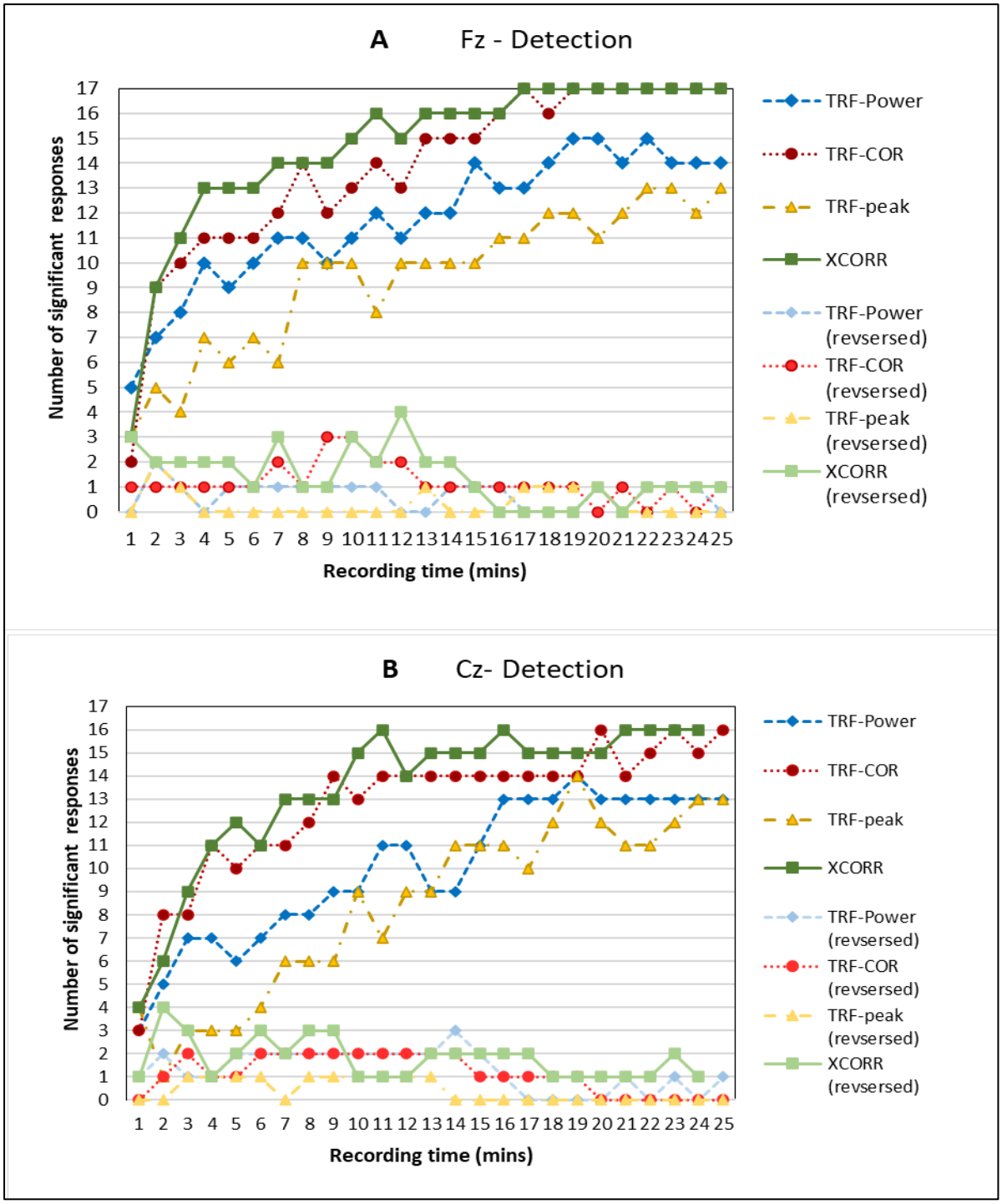

Figure 3.3 Number of subjects showing significant responses for each of the four parameters as a function of increasing EEG data. *A) EEG data from the Fz channel. B) EEG data from the Cz channel. Results are shown for forward speech and reverse speech (null hypothesis). For the latter, fewer than three detections were expected according to the 95% range of a binomial distribution of 17 measurements at*
$p < 0.05$.

For the time-reversed speech, any detections are considered false positives since no association was expected between the time-reversed envelope and the EEG. Figure 3.3 shows that these false positives were well controlled, remaining within the expected range of 0 - 3 significant responses, based on a binomial test for 17 tests with a 5% false positive rate. This

result was consistent in all cases, as shown in both panels A and B. Given the large number of tests conducted, a small number of exceptions are expected.

As the highest single channel sensitivity was found at Fz, this channel was then used to compare mean detection times; Figure 3.4 shows the mean detection times across subjects for the four parameters. Shapiro-Wilk testing revealed that the recording time data were not normally distributed ($p < 0.05$). A non-parametric Friedman test showed an overall significant difference between the parameters ($p < 0.05$). Additionally, Wilcoxon Signed Ranks tests for related samples were conducted across pairs of parameters. The mean detection time for XCOR at (4.82 min +/- 4.7 SD) was significantly lower than both the TRF-power (9.20 min +/- 9.6 SD) and TRF-peak (12.82 min +/- 9.7 SD) ($p < 0.05$ and $p < 0.01$, respectively). The TRF-COR mean detection time (6.41 min +/- 5.84 SD) was significantly lower than that of the TRF peak ($p < 0.01$). Figure 3.5 shows the distribution of the detection time for EEG data recorded from Fz for all subjects. A detection time of 26 min indicated that no response had been detected at 25 min into the recording. High variability in detection time across the subjects can be seen (e.g., XCOR showed detection after 2 min in subject 13, and 17 min in subject 17). In some subjects (e.g., 5 and 24), all analysis methods detected responses rapidly (e.g., by 3 minutes), while in other subjects (e.g., 11 and 12), the response was not significant even after 25 minutes.

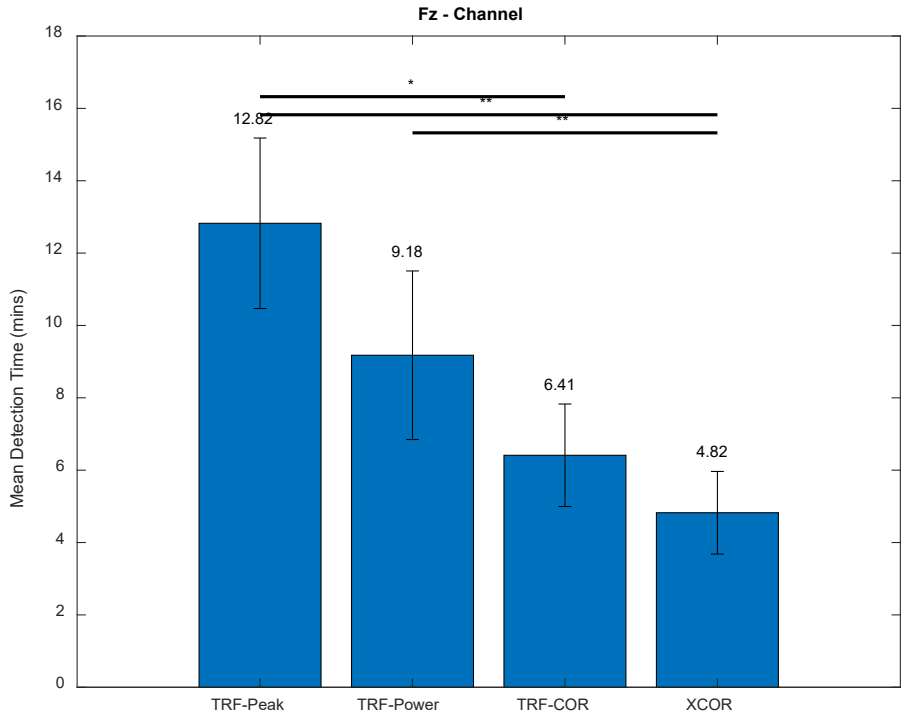

Figure 3.4 Mean detection times of the four testing parameters. *The figure shows the mean detection times of TRF-Peak, TRF-Power, TRF-COR, and XCOR at Fz. Error bars represent the standard error. Significant differences are indicated as follows: * = $p < 0.05$, ** = $p < 0.01$.*

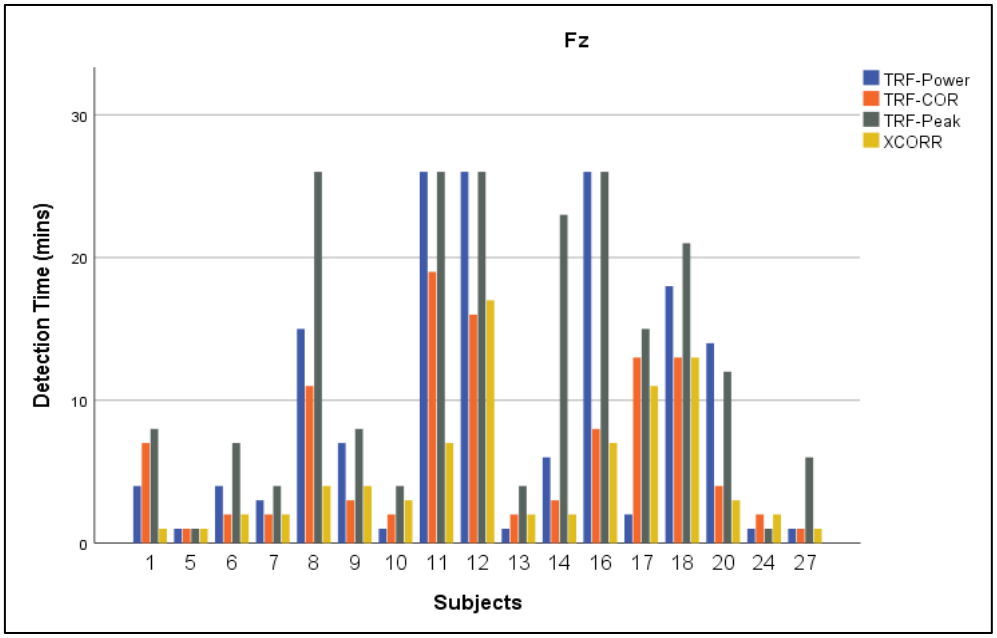

Figure 3.5 Individual detection time across subjects. *The figure shows the detection time for each subject across testing parameters TRF-power, TRF-COR, TRF-peak and XCOR at Fz.*

3.3.2 Detection times selecting the best-correlated channel from multiple channels

The analysis was repeated by analysing multichannel data and selecting the channel with the highest correlation value. Only the TRF-COR and XCOR parameters were used for this analysis because they exhibited the best performance in detection sensitivity (see Figure 3.3) and detection time (Figure 3.4). Figure 3.6 shows the detection rates of TRF-COR and XCOR when using one channel only (Fz), when selecting the channel with the highest correlation value from the set of 6 channels and when selecting the channel with the highest correlation value from all 30 channels. Due to long computational times, the results are displayed only for 5 min increases in time. As expected, detection rates increase with recording time for all methods. There was a decrease in sensitivity with using multichannel compared to a single channel, which was more pronounced when all 30 channels were included, compared to using only 6 channels in the frontocentral scalp region.

For the single-channel Fz analysis, 17 subjects show significant responses after 20 min using both parameters. When using 6 channels, XCOR detected responses in all subjects (n = 17) after 20 min, but TRF-COR did not reach 100% detection. Using all channels did not achieve detection in all subjects for either parameter. The maximum number of subjects in which detection was achieved using all channels was 16 for TRF-COR, and 15 for XCOR. Overall, multichannel analysis did not perform as well as single-channel analysis at Fz.

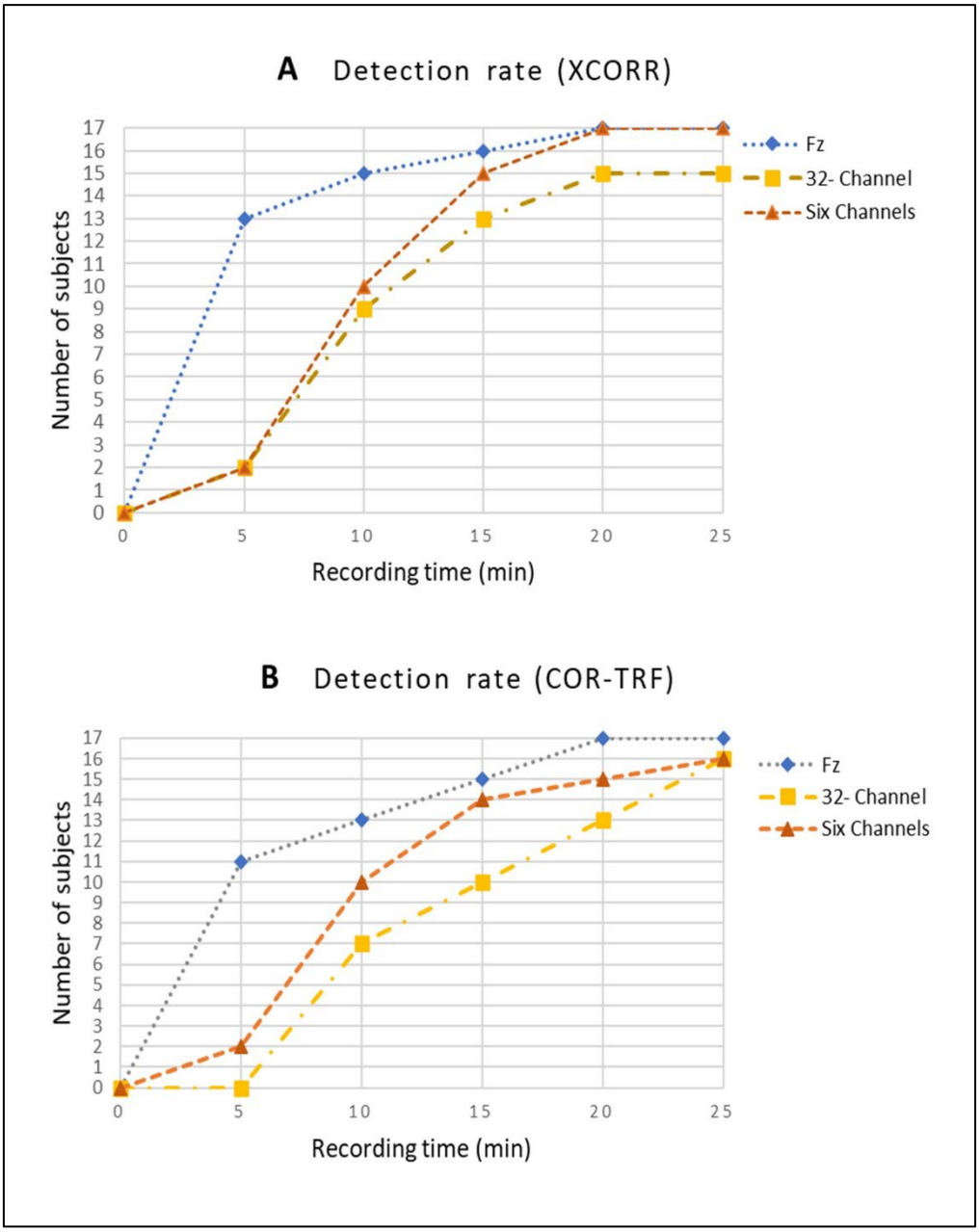

Figure 3.6 Detection rate comparing single, six and 32-channels. *The figure illustrates the detection rate using (A) XCOR and (B) TRF-COR parameters, analysed with three sets of EEG data: Fz only (dotted line), the best from 30-channels (dash-dotted line) and the best channel from six channels [Cz, Fz, F3, F4, FC1, and FC2] (dashed line). The figure shows the number of subjects showing detection with progressively increasing recording time in 5-minute steps.*

3.4 Discussion

Our main objective was to assess the feasibility of using EEG data from a single channel to detect cortical responses to continuous running speech. Two analysis approaches were compared to determine the presence of responses; cross-correlation and parameters obtained from the TRF analysis. Cortical responses were detectable in all subjects from the Fz channel using the XCOR and the TFR-COR parameters. The detection rate reached 100% after 17 min for both parameters. The mean detection time for XCOR of 4.82 min was numerically lower than

that of TRF-COR at 6.41 min, but the difference was not statistically significant in this relatively small sample. A post-hoc power analysis indicated that, for a difference of 1.59 min between XCOR and TRF-COR, 41 subjects would be needed to detect this with 80% power at $p < 0.05$. Our study was therefore underpowered to detect this relatively small difference; a larger study would be needed to explore this possible difference further. The performance of both parameters suggests good potential for single-channel clinical measurements. The cross-correlation parameter has an additional advantage that it is computationally less complex than the TRF-COR calculation. In contrast, response detection using TRF parameters TRF-peak and TRF-power failed to reach 100% detection in either channel even after 25 min, suggesting that they are less sensitive, with significantly elevated detection times compared to those of XCOR. It should be noted that the current study used a Biosemi system with active electrodes. Other systems with passive electrodes such as Interacoustics Eclipse (Interacoustics, 2022) may have different noise levels, and it may be useful to compare detection times on other systems in the future.

In the present work, the correlation coefficient values for XCOR and TRF-COR were low but consistent with what has been reported in previous literature: The XCOR correlation coefficient means were r = + 0.04 (maximum peak) SD (0.004) and r = -0.03 (minimum trough) SD (0.004). These findings are broadly comparable to correlation coefficient values reported by Kong et al. (2014) r = +0.07 and r = -0.03 . The difference in the positive value could be explained by the difference in the procedure as the signal processing is not identical in terms of the number of channels, recording time and filtering. In the TRF analysis, the mean correlation coefficient between the predicted and the actual speech envelope was r = 0.06 +/- 0.005 SD. Previous work by Di Liberto et al. (2015) and Di Liberto and Lalor (2017) showed similar correlations of r = 0.05 and r = 0.06, respectively.

When using the TRF method, the correlation coefficient between the predicted and actual responses is currently the most popular parameter for assessing model performance (Di Liberto et al., 2015, Di Liberto and Lalor, 2017, Kalashnikova et al., 2018, Jessen et al., 2019, Drennan and Lalor, 2019). Vanthornhout et al. (2019) found that the TRF filter displayed some similarities to those of CAEPs following short tone-bursts. Consequently, parameters derived from the TRF waveform, such as peak amplitude and latency, may provide additional insights into the neural responses. This is evidenced by attention effects on cortical tracking, which show a significant increase in TRF peak-to-peak amplitude when stimuli are attended to (Vanthornhout et al., 2019). In the present study, the detection sensitivity of the parameters TRF-peak and TRF-power was compared to the TRF-COR at the individual level: TRF-COR was able to detect responses in all subjects, whereas TRF-peak and TRF-power (which represent the shape of the TRF) showed non-significant responses in some subjects. The better detection observed with the prediction

accuracy of the TRF model may be attributed to the fact that the entire TRF is considered in the analysis. In contrast, the peaks in the TRF represent only specific and strong responses related to the stimulus. Although the TRF exhibits similar peaks to those seen in averaged AEPs in response to repeated short stimuli, it does not demonstrate sufficient detection sensitivity to be used as a standalone parameter.

Another explanation for the poor performance of single channel parameters that are based on the shape of the TRF might be related to the modelling of noise from frequencies irrelevant to the response. Since the impulse response attempts to model all frequencies in the input and output, frequencies between 30 Hz (the cut off frequency) and 64 Hz (half the sampling rate) contain little information regarding the input-output relationship, which may have impacted the reliability of the TRF-peak and TRF-power parameters. The issue may not be important when using TRF-COR as that parameter is not specifically affected by the shape of the TRF response.

Both the cross-correlation and the TRF approaches were successfully used to detect the response of the auditory system to continuous running speech for multichannel data (Lalor et al., 2009, Kong et al., 2014). It should be noted that the TRF and cross-correlation would be identical for white noise inputs (since the autocovariance matrix in Equation 2.2 would be an identity matrix). However, the TRF also accounts for the autocorrelation function (i.e., the spectrum) of the input signal (Crosse et al., 2016). Calculating the statistical significance of the cross-correlation peak is challenging using conventional statistical methods because: (1) the samples in each signal are correlated (i.e., the signals are not white noise); (2) the maximum cross-correlation value must be determined within specific response time lags and compared to the maximum cross-correlation of similar lags in noise. This makes it more complex than simply finding significance at each time-lag. In the present study, the effect of autocorrelation on the XCOR parameter test statistic was taken into account through bootstrap analysis. This method uses the original signals but applies random misalignment to remove the temporal correlation between the stimulus and the response, allowing us to estimate the null hypothesis for a given parameter.

The minimum time required to detect significant cortical responses was assessed previously by Di Liberto and Lalor (2017) using multichannel data. Using a subject-specific model, they found that at least 30 min of recording were needed to detect a significant level of phoneme activity across subjects. They used the correlation value between the predicted and actual EEG responses to assess the model's accuracy. Contrary to the outcome of Di Liberto and Lalor (2017), in the current study the time needed to detect a significant correlation (TRF-COR) in all subjects was 17 min, and in some subjects, responses were detected in 1 min (e.g., subjects 5 and 24) (see Figure 3.5). However, when comparing these results to Di Liberto and Lalor (2017),

some differences must be noted. First, the model used by Di Liberto and Lalor (2017) was built using specific phonemic speech features rather than the overall speech envelope, and individual phoneme responses may be harder to detect than the response to the speech envelope. Second, they averaged the correlation value across 12 channels in the fronto-central region, while the present analysis used a single channel. Averaging multiple channels meant Di Liberto and Lalor (2017) may have included weak responses, which may reduce sensitivity compared to single-channel analysis.

After comparing single channel detection at Cz and Fz, selecting the best channel from each subject (i.e., the channel with a maximum correlation coefficient between the EEG and the speech envelope) using either six central/frontal electrodes or 30 channels was explored. The aim was to measure any change in detection rate using the best channel compared to Fz. This approach comes with a high risk of selection bias, as choosing channels with the strongest response can increase the probability of false positives (Kilner, 2013). To avoid this, the statistical significance of each response was determined by estimating the bootstrap distribution across multiple channels. This selection approach may achieve higher correlation values; however, the multichannel bootstrap distribution's critical value is also expected to be higher than that of a single-channel case. As a result, a loss in sensitivity could conceivably occur with the multichannel method. Indeed, our results show a reduction in sensitivity when analysing responses across many channels and then selecting the best channel compared to using Fz alone. This reduction in sensitivity with multiple channels was most evident when all channels were included, compared to only six pre-selected channels. This suggests that using fewer channels can improve sensitivity because noisy channels without responses are removed. Montoya-Martínez et al. (2021) showed that by reducing the number of channels from 64 to 20 in a backward model, the correlation between the reconstructed and actual speech envelopes improved substantially. Of course, this can only be expected if the best channels are retained.

One consideration in the current analysis is the method of validation used. Most previous studies used different sets of EEG data to train and test the model for cross-validation (O'Sullivan et al., 2015, Di Liberto and Lalor, 2017, Vanthornhout et al., 2019). The ‘leave one out’ method is commonly applied, where roughly 80 % of the data is used to train the model and the remaining 20% to test it (Crosse et al., 2016). However, as one of the present study's analysis objectives was to investigate the effect of time (every minute) on response detection, this approach was not appropriate. It is also debatable if cross-validation is needed when simply detecting responses. Therefore, the analysis in the current study did not include the cross-validation approach. Instead, the TRF parameters were trained and tested on the same dataset to determine whether they were significant compared to the TRFs of non-aligned data.

Whilst using the same data for training and testing could result in bias and higher correlation values than the 'leave one out' method without bootstrapping, by testing the significance using the bootstrap method, critical values used compensate for the potentially biased estimator and hence false positives are avoided. The low number of false positives (within the expected range) provides support for the approach taken here. To our knowledge, this approach has not been used previously for cortical responses to speech.

The methods used to analyse responses to continuous speech might potentially be improved in future work: the forward model performance using one channel could be compared to the backward model. This will assess whether the backward model that uses multiple channels can improve detection sensitivity. Other potential developments of methods include limiting the frequency range used to record the TRF-peak parameter (excluding estimation of the transfer function at frequencies not in the response) and focusing the analysis of cross-correlation parameters to more specific time lags where responses are expected to occur.

3.5 Conclusion

This chapter explored the detection of cortical responses to running speech using a single channel. Most previous studies in this area have only used multichannel data. The findings here suggest that detection with single channel data is possible. Significant responses were detected in all subjects (N = 17) using the Fz channel with either the cross-correlation parameter XCOR or the TRF correlation measure TRF-COR. Numerically the XCOR method appears most sensitive with a mean detection time of 4.8 min compared to 6.4 min for TRF-COR, but this difference did not reach statistical significance due to the small sample size. Detection times for the parameters TRF-peak value and TRF-power were significantly higher.

The approach of selecting the channel with the highest correlation from multiple channels resulted in a reduction in sensitivity compared to single-channel analysis, most likely due to increases in critical values when using multiple channels. In summary, these results are promising for clinical applications using single-channel EEG recordings to detect cortical responses to continuous running speech. However, this chapter only included intelligible speech with participants paying attention, which may not be representative of target clinical groups such as infants. Future studies will explore these factors using the same single-channel analysis approach.

Chapter 4 Comparing Cortical Responses to Continuous Speech and Speech-Modulated Noise During Passive Listening

Building on the high detectability of single-channel EEG observed in Chapter 3, the experiment described in this chapter was designed to investigate the effects of attention and stimulus intelligibility on the single-channel detection of cortical responses to continuous speech. This study was part of a collaborative experiment conducted with Suwijak Deoisres, who was also a PhD candidate. Because COVID-19 disrupted and delayed plans for initial experimental work, the decision was made to run a joint experiment. Suwijak set up the stimuli in Praat (an open-source software used for acoustic analysis; (Boersma and Van Heuven, 2001), prepared the BioSemi for EEG, and created the poster for participant recruitment. Suwijak and I performed participant recruitment and data collection jointly. Suwijak also helped with editing the code for the EEG data analysis. I carried out the EEG data analysis and statistical analysis in this chapter.

4.1 Introduction

As highlighted earlier in the project, the primary goal is to develop a clinically viable tool to objectively test speech perception in infants. One component involves measuring responses without requiring active attention to stimuli and using a single EEG channel. Another important factor is the type of speech stimulus. Utilising a stimulus that is acoustically similar to speech but unintelligible allows the creation of a tool that is universally applicable, regardless of the subject's native language. Therefore, the current study will explore the possibility of detecting cortical responses to continuous speech during passive listening and will examine the effect of speech intelligibility by comparing responses to both intelligible and unintelligible speech.

4.1.1 Effect of attention and passive listening.

It is well-established that attending to a stimulus enhances cortical responses to sensory input. However, it remains uncertain whether passive listening will lead to a significant detection of cortical responses at the individual level. Using a cross-correlation analysis approach, Kong et al. (2014) investigated the effect of attention on the neural tracking of the speech envelope when subjects were presented with an additional sensory stimulus (visual). The stimulus was presented either with a silent movie (cross-modality condition) or a competing sound (within-modality condition). The movie condition was assigned as the quiet condition as subjects were

instructed to either attend to or ignore the stimulus presented while watching. For the competing listening condition, subjects were asked to attend to one of the sounds. The study included eight normal-hearing adults. In the quiet condition, the cross-correlation function between the EEG and speech envelope showed no significant difference between active and passive listening. However, enhanced N1 and P1 peaks were indicated in response to the attended sound in the competing-speakers conditions. Kong et al. (2014) concluded that top-down attention affects the cortical neural tracking of the speech envelope and that this effect depends on the task because it was not indicated in the quiet condition.

As mentioned in Chapter 2 (see Section 2.4.3), Vanthornhout et al. (2019) examined the differences in neural tracking of speech envelopes during active and passive listening. Their findings indicate that while attention can enhance the strength of cortical responses, its effect on neural tracking of the speech envelope is only evident in the presence of noise. Since they did not observe an attention effect at a high SNR, the findings of Kong et al. (2014) and Vanthornhout et al. (2019) are consistent. However, both studies examined the attention effect at a group average level. To our knowledge, no previous research has explored detection sensitivity in passive listening at an individual level. When testing infants, it would be useful to detect significant cortical responses to speech while they are passively listening to the stimuli and engaged in watching something. Therefore, this study will measure the detectability of speech during passive listening.

4.1.2 Effect of speech intelligibility

When looking into the speech stimulus, the slow modulation of the envelope is one of the essential features of speech intelligibility (Drullman et al., 1994, Shannon et al., 1995). A number of existing studies have demonstrated the possibility of cortical tracking of the low-frequency natural speech envelope (Aiken and Picton, 2008a, Lalor and Foxe, 2010, Ding et al., 2014, Di Liberto et al., 2015). In addition, neural responses to the stimulus envelope were detected even when the speech was unintelligible (Howard and Poeppel, 2010, Zoefel and VanRullen, 2016). It would be advantageous to test various infant groups using a universal signal that is not language specific, which allows it to be used for testing speech perception independent of subjective native language. However, it is still unclear whether using unintelligible speech or a speech-like stimulus will affect the cortical neural tracking of the speech envelope.

Etard and Reichenbach (2019) investigated how cortical responses reflect the comprehension and clarity of speech by recording EEG responses in native and foreign languages. EEG data were recorded from ten normal-hearing native English speakers while they listened to

continuous English and Dutch speech in quiet conditions and with various SNR conditions. No difference was shown between the decoder reconstruction accuracy of the continuous English and Dutch speech presented to English speakers in the quiet condition. With background noise, the neural entrainment of the speech envelope was significantly higher with English speech only in the delta frequency band. Based on the results of this study, it appears that comprehension of the stimulus is not essential in detecting significant cortical tracking of the stimulus envelope. Significant cortical responses can be seen when incoherent stimuli with similar auditory characteristics to those of intelligible speech are presented.

Zou et al. (2019) examined how different auditory and linguistic processes affect envelope-tracking speech responses in noisy settings. The neural tracking of the speech envelope was compared between two groups of listeners, i.e., foreign listeners who do not comprehend the test language and native listeners of the test language. In this study, 32 normal-hearing adults participated; 16 were native Mandarin speakers without prior exposure to Cantonese, and the remaining 16 were native Cantonese speakers. The study revealed that foreign listeners' neural reconstruction accuracy was more sensitive to noise than native listeners. The foreign listers group showed significantly higher reconstruction accuracy with +9 and -6 SNR levels. The inconsistent findings between Etard and Reichenbach (2019) and Zou et al. (2019) illustrate how the EEG response does not always provide a definitive explanation of what aspect of speech perception has been evaluated. However, the difference between comprehensible and incomprehensible speech regarding cortical detectability could be further examined by exploring other factors, such as detection time and sensitivity. Having a stimulus that can evoke significant cortical responses, which is acoustically similar to speech yet not language-specific, will be important for implementing a universal speech test in infants.

The current study tries to answer the following questions: Can significant cortical responses to continuous speech be detected using data from one channel under passive listening conditions?; Is there a significant difference in detection times and rates between cortical responses to natural English speech and speech-modulated noise recorded for native English speakers?

Cortical responses were collected while subjects listened to stimuli passively while watching a silent movie. The subjects' native language and non-intelligible speech-modulated noise were used to elicit cortical responses. EEG responses to stimuli were recorded for normal-hearing subjects. Both cross-correlation and the TRF forward model were applied to detect the responses from one channel.

4.2 Materials and Methods

The experiment involves conducting a hearing screening test and EEG recording.

4.2.1 Participants

Twenty-two adults with normal peripheral hearing thresholds who were aged 18–40 participated. All participants were native English speakers. To ensure that the hearing sensitivity was within normal limits, screening pure tone audiometry was performed for each participant.

4.2.2 EEG recording

The EEG data was collected using a BioSemi ActiveTwo recording system. During the EEG recording, the participant was seated in a comfortable chair and asked to focus on the screen while ignoring the stimulus. A series of silent documentaries with English subtitles were presented via a TV screen placed in front of the participants. See Figure 4.1 for the experiment setup.

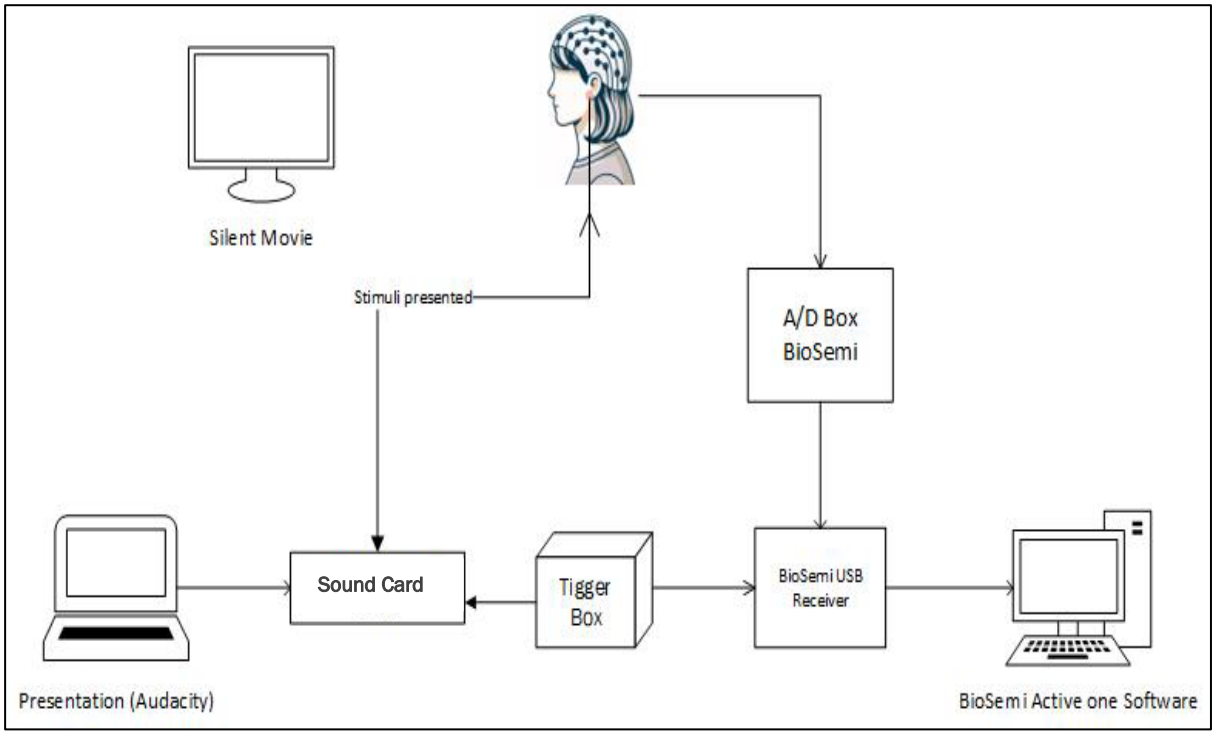

Figure 4.1 The diagram illustrates the EEG experimental setup. *The auditory stimuli are processed and presented through a RME Fireface UC audio interface, controlled by a computer running Audacity software. The Trigger Box serves as a central point of integration, sending precise signals to synchronise the auditory stimuli with the data collection process. The data acquisition is conducted through an A/D Box from BioSemi, which converts the analog signals from the participant into digital data. This data is received by a BioSemi USB Receiver connected to a separate computer equipped with BioSemi ActiView software. A silent movie is displayed on a monitor to control participant attention.*

4.2.3 Stimulus

This experiment has two conditions, each with a different stimulus. The first was a continuous speech from Chapter 4 of *Children of Odin* by Pádraic Colum, an audiobook read by a female narrator (available to download for free on https://librivox.org/the-children-of-odin-by-padraic-colum/). The second was speech-modulated noise. For comparability, the modulated noise was created to be as similar to speech as possible in terms of its power spectrum (speech-shaped noise). The spectral shape was stationary, so only the amplitude was modulated. The two stimuli were similar in envelope intensity and long-term (average) spectrum. Each stimulus was presented in four blocks, with each lasting for about 3 min 45 s (total stimulus length of about 15 minutes). The sampling rate of the presented stimuli was 44.1 kHz, which was up-sampled from 22.05 kHz (the original sampling rate of the mp3 file). The stimulus segments were presented monorally to the right ear at 65 dBA LeqA SPL through Etymotic ER-2A insert earphones (Etymotic Research, Inc., Illinois, USA).

The speech-modulated noise stimulus was generated using standard functions in the Praat software (Boersma and Van Heuven, 2001). A speech-shaped noise, the carrier signal, was created by generating Gaussian white noise filtered by the long-term average spectrum of all continuous speech stimuli used in this study concatenated, with a bandwidth of 0 Hz –12 kHz. The speech-envelope-modulated noise was then generated by multiplying the speech-shaped noise with the amplitude envelope of the continuous speech, which served as the modulating signal. The amplitude envelope of the continuous speech was extracted using Praat software. This process involved converting the speech .wav file to an intensity tier, which facilitated the envelope extraction. Specifically, the squared samples in the .wav file were processed using a Gaussian analysis window (weighted averaging) to obtain a smoothed envelope. Figure 4.2 shows a segment of the envelope of the continuous speech and speech-modulated noise intensity envelope. The Pearson correlation coefficient of $r = 0.974$, averaged across all stimulus segments, indicates a strong relationship between the two envelopes.

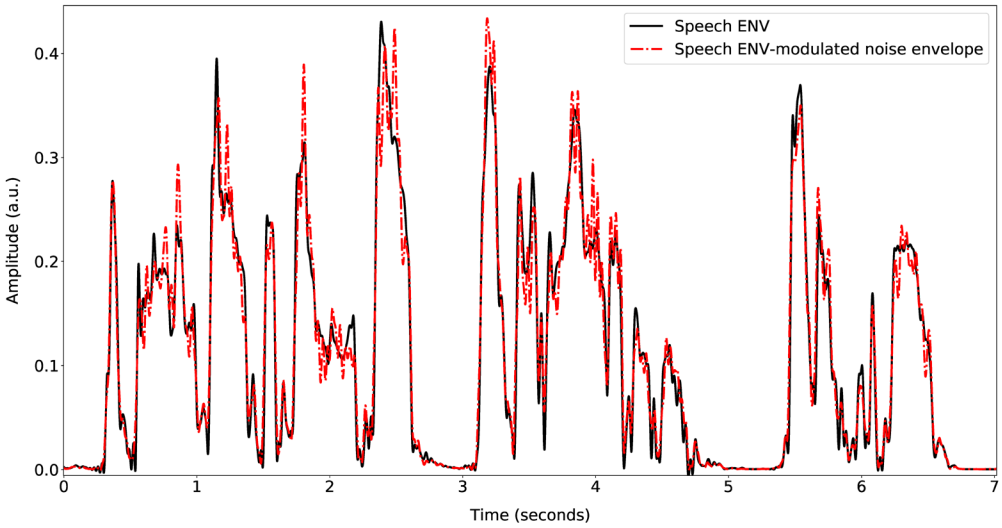

Figure 4.2 Temporal envelopes of speech and speech-modulated noise*.This figure provides a comparison between the natural variations in the amplitude of a speech signal and the corresponding modulations in the speech-modulated noise. The black line represents the speech envelope, while the dotted red line represents the speech-modulated noise envelope.*

4.2.4 Data analysis

Data analysis took place using MATLAB 2019a. The EEG responses were recorded from 32 channels, down-sampled to 128 Hz and referenced to the two mastoid electrodes. The EEG was filtered between 1 Hz and 30 Hz, covering the same band as the envelope for the speech signal. The EEG data were analysed from a single fixed channel, either the vertex (Cz) or the high forehead (Fz). The recorded EEG data were analysed using two primary analyses: the cross-correlation of the speech envelope with the EEG signal and the TRF forward model.

A detailed explanation of the cross-correlation and TRF analysis methods was provided in Chapter 2 (see Section 3.2.1). However, for the current analysis, only one parameter of the TRF was used, namely TRF-COR. As shown in the previous chapter, this approach had the highest detection sensitivity out of TRF detection approaches (see Figure 3.3). False positive rates were also tested to verify that the proposed analysis methods were working as expected.

4.2.5 Detection time analysis

The minimum detection time required for the response was recorded for each subject. The purpose of this was to compare the two stimulus conditions (speech and speech-modulated noise). The data were divided into segments of increasing length from 1 min to 15 min in 1 min increments. For each segment, the signal detection analysis outlined above for each parameter

was used to indicate whether a response was present. The detection time determined for each subject was the minimum recording period after which a significant response was detected. In cases in which a subject showed no detection, the detection time was set to 20 min (above the maximum recorded duration) to avoid missing data in the statistical analysis. This approach ensured that all cases, whether with or without detection, were included in the analysis of detection times.

4.3 Results

Two analysis approaches were applied to detect the cortical responses to continuous speech: the TRF forward model correlation and the maximum cross-correlation between speech envelope and EEG. The two approaches were applied using EEG data from one channel (only Cz or only Fz) and both continuous speech and speech-modulated noise were tested.

4.3.1 Detection rate

Table 4.1 shows the measured false positive rate for each of the parameters. Tests were applied to the two parameters (TRF-COR and XCOR) using white noise as the input and output signals. In each case, 1000 iterations were tested. All the methods produced a false positive rate close to the expected value of 5%, or 50 out of 1000 false positives (within the 95% range of detection rate values expected from the binomial distribution when using a test at $p < 0.05$ with 1000 repetitions, or between 37 and 64 false positives).

Table 4.1 Analysis of false positive (FP) rates for analysis parameters using white noise as input and output signals. *The table shows the expected FP range for 1000 tests was calculated when testing at p < 0.05 from the binomial distribution.*

Parameter	TRF-COR	XCOR
#Tests	1000	1000
FPs	46	53
% FP rate	0.046	0.053
Expected FP Range	37–64	37–64

Figure 4.3 shows the number of subjects with significant detection with progressively increasing amounts of data by recording time. The results from the speech stimulus condition indicate that after 15 minutes, the detection rates were 63.6% (14 out of 22) with XCOR and 59% (13 out of 22) with TRF-COR for the Fz channel. Additionally, the detection rates were 68% (15 out of 22) with XCOR and 59% (13 out of 22) with TRF-COR for the Cz channel. The results for the modulated noise stimulus show that after 15 minutes, the detection rates were 77% (17 out of

22) with both parameters for the Fz channel. In contrast, the detection rates were 86% (19 out of 22) with XCOR and 77% (17 out of 22) with TRF-COR for the Cz channel.

McNemar tests were run to determine whether there were statistically significant differences in detection rate after 15 min between the two stimulus conditions, the testing parameters and the electrode locations. The results showed no significant difference between the stimuli across all testing conditions. Also, no significant difference was indicated between the two testing parameters or the two electrode locations. The detection sensitivity was generally higher for the modulated noise than speech, but this difference was not significant. Figure 4.3 also shows the false positives for both stimuli, (A) speech and (B) modulated noise, using the reversed speech envelope as an effective no-stimulus condition. All subjects were analysed, with the range of false positive rates from the binomial distribution under a no-stimuli condition predicted between 0 and 3. False positive rates for both stimuli across all testing conditions were within the acceptable range.

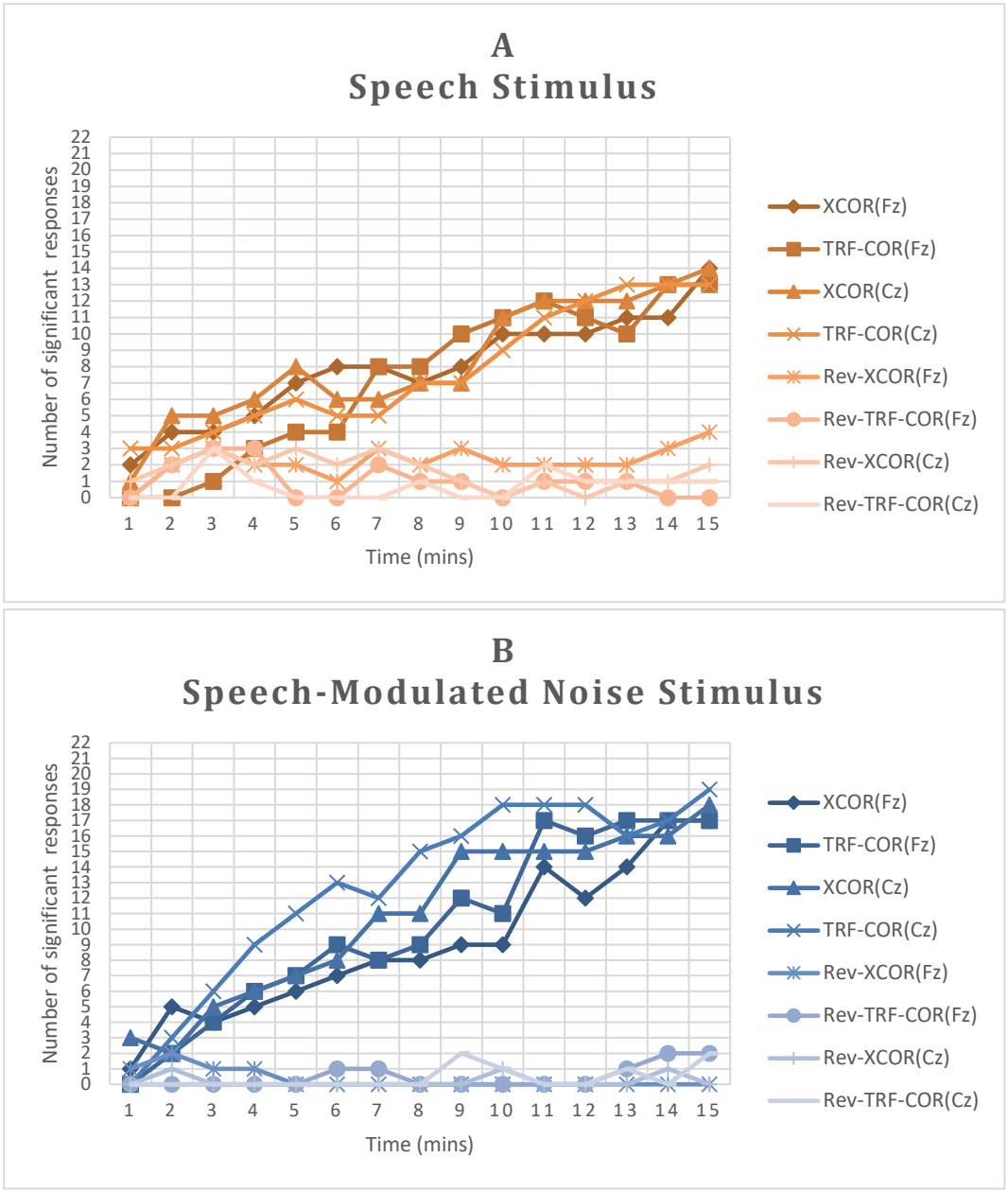

Figure 4.3 Number of subjects showing significant responses for each parameter used in the analysis as a function of increasing EEG data. *The figure illustrates the number of subjects showing significant cortical responses for each parameter used in the analysis (TRF-COR and XCOR) with progressively increasing amounts of data. The figure is divided into two parts: (A) shows the detection results with the speech stimulus, while (B) shows the results with the speech-modulated noise stimulus. The dark-coloured lines represent the results with the forward stimuli analysis, while the light-coloured lines represent the results with the reverse stimuli analysis, which serves as a control to determine the false positive rate.*

Figure 4.4 shows the detection rate for all testing conditions, electrode configurations and analysis methods. In general, detection rates did not reach 100% for any of the passive listening conditions used in this experiment. The mean detection rate for the modulated noise condition (80.68%) is higher than that for the speech condition (61.36%).

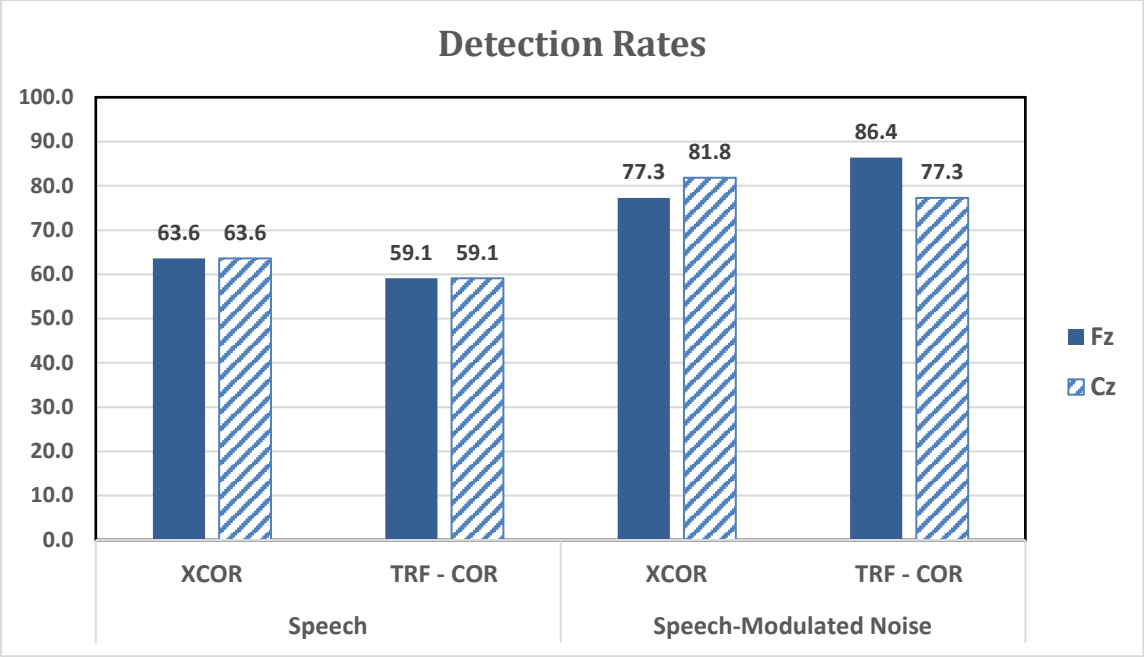

Figure 4.4 Detection Rates for speech and speech-modulated noise with passive listening
Detection rates were obtained using the XCOR and TRF methods of analysis for data from the Cz (blue patterned bars) and Fz (blue solid bars) and a stimulus duration of 15 min. The left four bars display the results for the speech stimulus: the first two bars show the XCOR outcomes, and the next two bars show the TRF outcomes. The right four bars display the results for the speech-modulated noise, following the same order as the speech results, with the first two bars showing the XCOR outcomes and the next two bars showing the TRF outcomes.

4.3.2 Detection time

In addition, the mean detection time for the speech and the speech-modulated noise conditions were compared. As mentioned above (see Section 4.2.5), all subjects were included in the analysis, even those without detection. The two conditions were compared using both analysis parameters (XCOR and TRF-COR). The comparison was also made using two selected EEG channels (Fz and Cz). Shapiro-Wilk testing revealed that the test time data were not normally distributed ($p < 0.05$), so it was necessary to use non-parametric testing to search for significant differences in test times between conditions.

Wilcoxon signed-ranks tests for related samples were conducted to compare the speech and modulated noise stimuli given the same parameters and electrode locations. Figure 4.5 (Right Panel) illustrates the mean detection times for the Fz channel. The detection time for speech-modulated noise appears to be slightly shorter than that for the speech stimulus for both parameters; however, the statistical test showed no significant difference. Figure 4.5 (Left Panel) indicates the mean detection time for the Cz channel for both detection parameters. With TRF-COR, the detection time for the modulated noise stimulus (8.18 min +/- 6.2 SD) was significantly lower than that for the speech stimulus (12.68 min +/- 6.7 SD), with $p < 0.05$. No

significant difference was found between stimuli with the XCOR at Cz (p = 0.068). Overall, there is a trend for the noise-evoked response to produce a slightly lower detection time than the speech-evoked response, but this was only significant for the TRF-COR parameter at Cz.

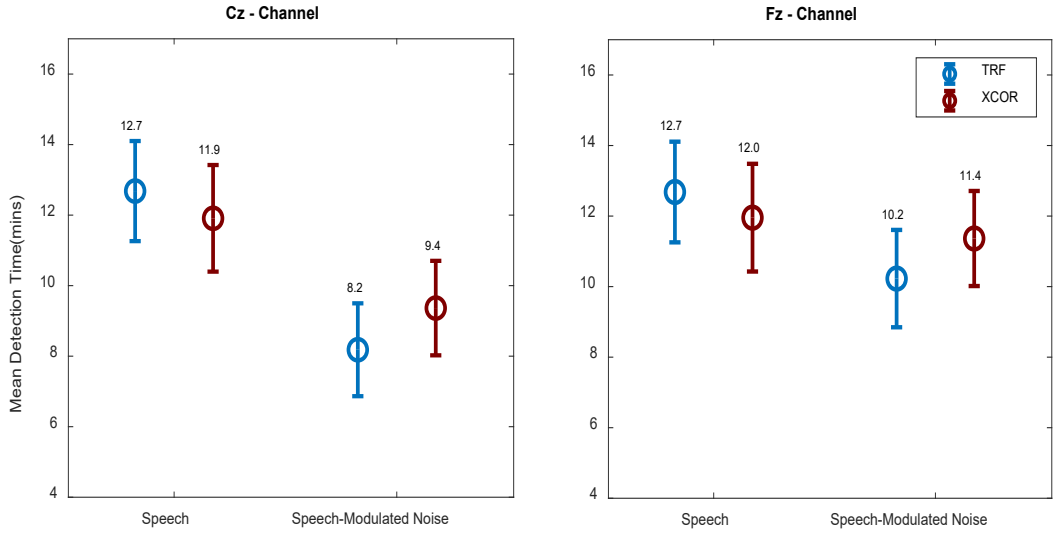

Figure 4.5 Mean detection times for speech and speech-modulated noise stimuli at channels Fz and Cz. *The error bars indicate the SE of the mean. The y-axis represents the detection time in minutes. The blue bars show the result of using the TRF method of analysis, while the maroon bars show the results of using XCOR. The L box shows the results for Cz, while the R box shows the results for Fz.*

Additional comparisons were made to investigate the effect of stimulus type and channel location on detection time. First, the results for detection time by averaging the outcome of the two channels (Fz and Cz) were compared. Wilcoxon signed-ranks tests showed that with the TRF-COR parameter, the detection time for the modulated noise stimulus was significantly shorter (9.2 min ± 5.35 SD) than for speech (12.68 min ± 6.05 SD), with $p < 0.05$. No significant difference was found with XCOR, where $p = 0.26$. Second, the two-channel locations were compared by averaging the detection time for the two stimulus types. No significant difference was found between the EEG channels for either detection parameter (XCOR: $p = 0.29$ and TRF-COR: $p = 0.17$).

4.3.3 TRF correlation values

The TRF correlation values (reflecting the accuracy of the forward model) were used to further evaluate the differences between the two stimulus conditions and channel locations. Figure 4.6 shows the mean correlation values of the modulated noise and speech stimulus in two different channel locations (Fz and Cz). Based on the Kolmogorov-Smirnov normality test, there is no significant evidence to suggest that the correlation values deviate from a normal distribution ($p > 0.05$). A two-way repeated measures ANOVA was performed to compare the effect of

stimulus type and channel location on correlation values. There was a significant main effect of stimulus type on correlation values ($F[1,21] = 7.2$, $p < 0.05$). The correlation value of the TRF was higher with the modulated noise stimulus (mean = 0.044 +/- 0.002 SD) as compared to the speech stimulus (mean = 0.038 +/- 0.002 SD). There was no significant main effect on the part of channel location on correlation values ($F[1,21] = 0.292$, $p > 0.05$). Correlation values were similar when using Fz (mean = 0.042) or Cz (mean = 0.041). There was no significant interaction between stimulus type and location ($F(1,21) = 0.102$, $p > 0.05$).

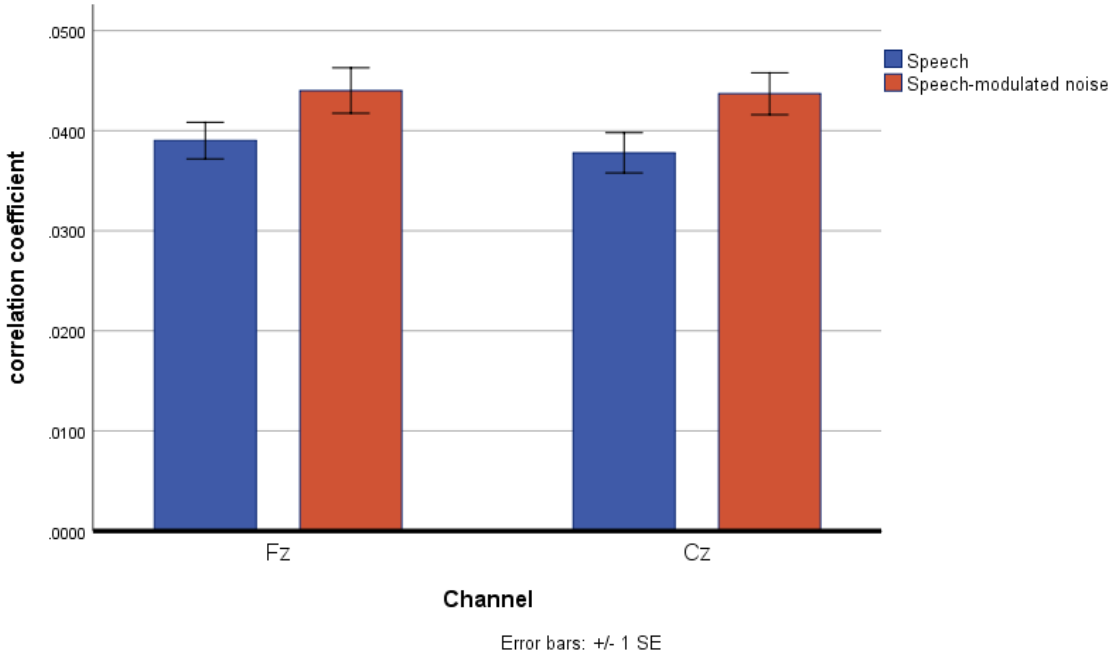

Figure 4.6 Correlation between predicted and actual speech envelope for TRF model comparing speech and speech-modulated noise stimuli. *The figure shows the correlations measured from the Fz channel (left two bars) and the Cz channel (right two bars). The error bars represent the mean and standard error (SE). The blue-coloured bars correspond to the results for the speech stimulus, while the orange-coloured bars correspond to the results for the speech-modulated noise stimulus.*

The TRF filter waveform represent the estimated impulse responses based on the forward modelling. Figure 4.7show the grand average TRF across subjects, highlighting the differences between the two stimuli used in the study.

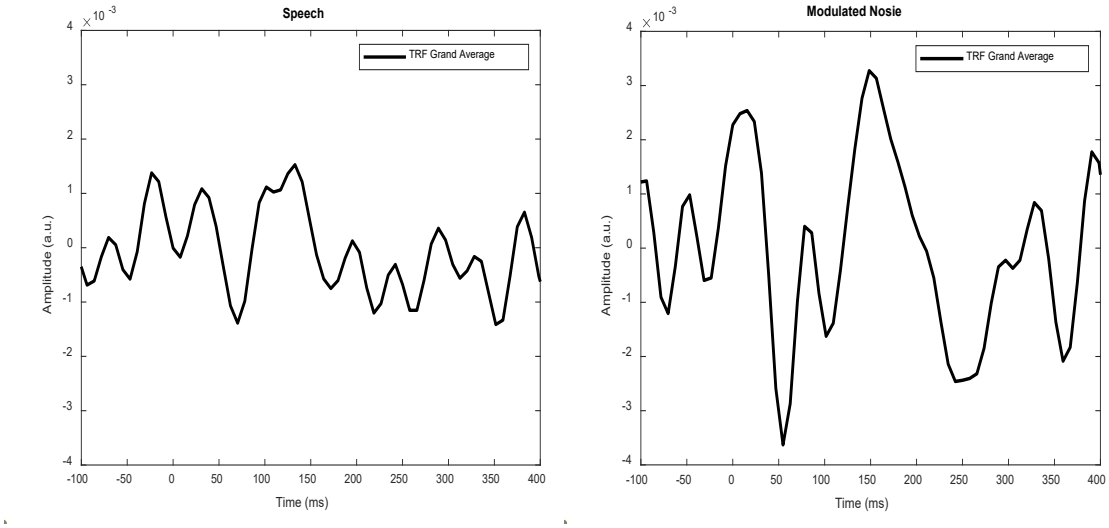

Figure 4.7 Grand average of the TRF filters from the Cz channel for speech and speech modulated noise*. The left panel shows the averaged TRFs for the speech stimulus condition, while the right panel shows the averaged TRFs for the speech-modulated noise stimulus condition.*

4.3.4 XCOR function

The XCOR function represents the cross-correlation between the stimulus envelope and the EEG. The group-averaged XCOR function was visually compared between the two stimuli. Figure 4.8 shows the XCOR function averaged across subjects for both the speech and speech-modulated noise stimuli. It is evident from the figure that the speech-modulated noise has higher peak amplitudes compared to speech.

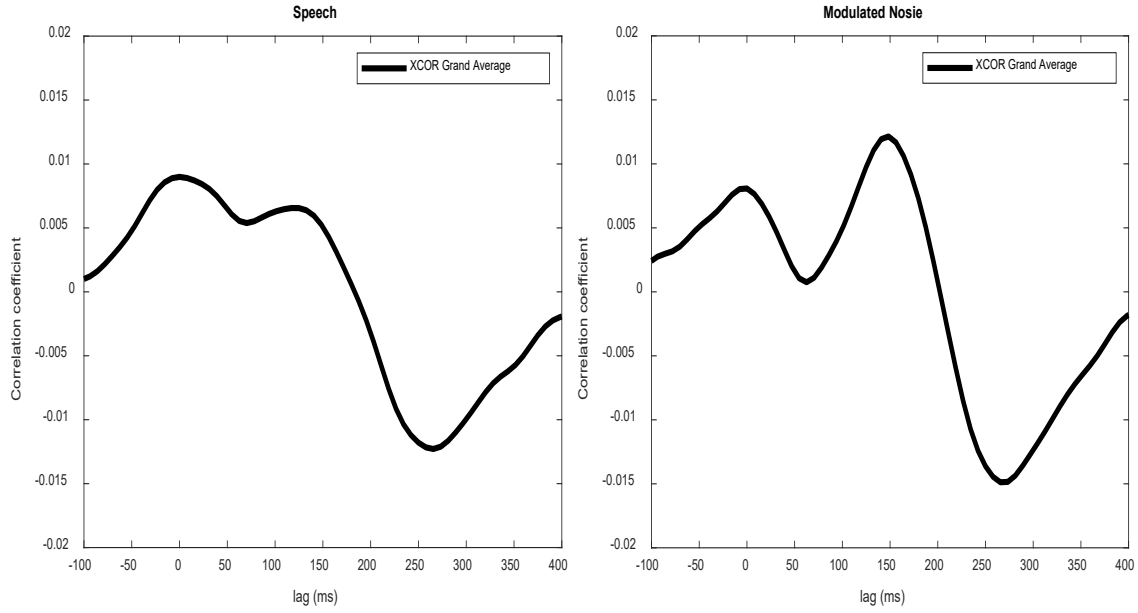

Figure 4.8 Grand average of the cross-correlation function (XCOR) from the Cz channel for speech and speech modulated noise*. The left panel shows the cross-correlation*

function for the speech stimulus condition, while the right panel shows the cross-correlation function for the speech-modulated noise stimulus condition.

4.3.5 EEG signal variability

The EEG data recording captures the responses to the specific stimulus presented and includes a wide range of brain activity and potentially other strong additive responses. This can lead to increased variability in the signals, which may affect the response of interest. One method of estimating this variability is calculating the recorded signal's standard deviation. It is important to note that the SD reflects contributions from both the true AEP and noise; however, the AEP contribution to the SD was assumed to be negligible compared to the noise. This analysis calculated the standard deviation of the responses for both channels.

Figure 4.9 shows the average standard deviation of the EEG recorded from the Fz and Cz channels for both stimulus conditions. For each channel, a total of 22 SD values were included, corresponding to the number of subjects tested. Kolmogorov-Smirnov and Shapiro-Wilk tests suggest that the EEG noise level is not distributed normally ($p < 0.05$). The Friedman test was performed to compare the effects of stimulus type and channel location on noise level. The test showed a significant difference between the noise level measured with different channel locations and stimulus types ($\chi^2(df = 3) = 38.45$, $p < 0.001$). A post-hoc test using a Wilcoxon single-rank test showed no difference between the noise levels of the speech-modulated noise and speech stimuli for the Fz and Cz channels. Significant differences were found between the channel locations for both stimulus types ($p < 0.001$), with Fz showing a higher SD.

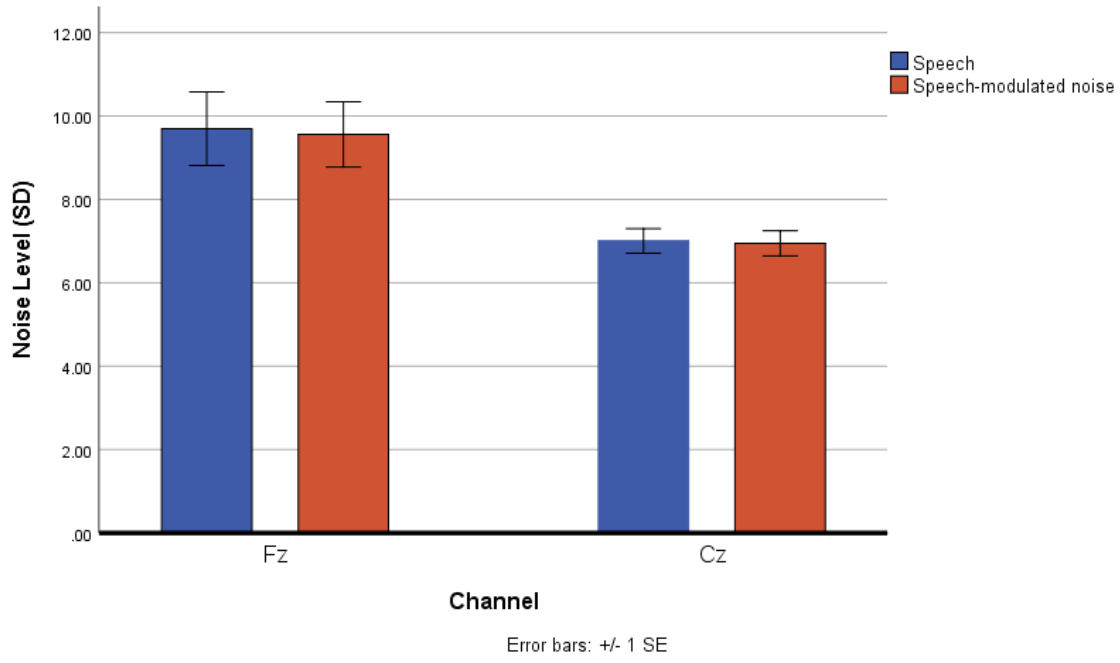

Figure 4.9 Mean Noise level of EEG data (standard deviation) for Fz and Cz channels under speech and speech-modulated noise conditions. *The figure shows the mean of the noise level measured from the Fz channel (left two bars) and the Cz channel*

(right two bars). The error bars represent the mean and standard error (SE). The blue-coloured bars correspond to the results for the speech stimulus, while the orange-coloured bars correspond to the results for the speech-modulated noise stimulus.

The correlation between SD value and detection time and rate was measured to investigate whether the variability level in the EEG affects the measured responses. The SD of the recordings where responses were detected was compared to those where responses were not detected for the Cz channel using the XCOR detection results. Mann-Whitney U-tests showed no significant difference in noise levels for either stimulus type: modulated noise ($U = 23$, $n = 22$, $p > 0.05$) and speech ($U = 44$, $n = 22$, $p > 0.05$).

A Spearman correlation coefficient was computed to determine the relationship between variability level and detection time. This correlation was measured for the two channels in each condition and analysis method. A significant relationship between the variables was observed only in the speech condition results for XCOR in both channels (Fz: $r = 0.455$; Cz: $r = 0.424$) and for TRF in the Fz channel ($r = 0.481$), with $p < 0.05$. However, after applying a Bonferroni correction for multiple comparisons, these results were no longer significant (corrected p-value threshold: 0.012, with four pairwise comparisons per channel).

4.4 Discussion

This study investigated the possibility of using passive listening and a universal speech stimulus in measuring cortical responses to speech in normal-hearing English-speaking adults. It also explores the use of two stimuli, one which was language-specific, natural English speech, and one which was non-language-specific: speech-modulated noise. Cortical responses were analysed using TRF and XCOR approaches via two single channels (Cz and Fz). The two analysis methods broadly produced similar findings. It appears that, when using passive listening a response is not detectable in all subjects for either stimuli and 15 min recording. Maximum detection was measured at the Fz channel, with higher detection sensitivity when using speech-modulated noise. The mean detection time for modulated noise was consistently lower for all testing conditions than when using the speech stimulus. However, it was only significantly lower when measured from the Cz channel and analysed using TRF-COR. The prediction accuracy of the TRF model was higher for the modulated noise stimulus than for the speech stimulus. In general and in most subjects, the speech-modulated noise stimulus analysed with TRF showed a robust cortical outcome measure in terms of the detection time and prediction accuracy of the TRF model.

In the current study, the detection rate using a single EEG channel failed to reach 100%, unlike in the previous study (see Figure 3.3). One major difference between the two studies was the attention task methodology. In the earlier study, subjects were asked to attend to a stimulus, while here, they were asked to attend to a silent documentary with subtitles and ignore the stimulus. Previous studies have provided evidence for more robust neural tracking of the speech envelope in the presence of active attention (Vanthornhout et al., 2019, Alickovic et al., 2020). Thus, the reduced detection of neural responses with passive attention in the current study compared to the previous one appears consistent with those findings. Another key difference between the current and previous studies (Chapter 3) was the recording duration. In the current study, only 15 minutes of EEG recording was conducted in each condition, compared to 25 minutes in the previous study. In Chapter 3, detection reached 100% after 17 minutes of recording. In the present study, a shorter recording time was chosen to assess the feasibility of detection within a timeframe appropriate for infant testing. Also, the quiet conditions used in this analysis were part of an experiment used for a different PhD project that included other testing conditions (Deoisres, 2023). The total test time for the experiment was about 3 hr. Therefore, it was only feasible to have 15 min to test responses in the two quiet or no noise conditions included in this study.

One cause of low detection sensitivity may be the quality of the EEG recording. One way to describe EEG noise is by measuring the variability distributed around the mean of zero, which can be quantified by the SD (Hall, 2015). Therefore, the SD was used to measure the variability in the EEG signal and its relation to response detectability. The results show no significant difference when comparing the mean noise levels between the detected and non-detected groups. The only correlation between detection time and noise level was for the Fz channel. As mentioned above, a detection time of 20 min was assigned to subjects without cortical responses. These higher detection time values may have introduced bias into the measured significant positive correlation. Although significant correlations were observed in some conditions, the correlations were not very strong, as they did not remain significant after correction. Additionally, removing outliers (non-detectable subjects) resulted in no significant correlations. The results suggest that the standard deviation as a measure of EEG signal variability or noise level does not effectively reflect the detectability of cortical responses to continuous speech when using single-channel analysis.

To our knowledge, this is the first study that has measured cortical responses to continuous speech with passive listening at an individual level. Previous studies that investigated the effect of attention on the neural tracking of the speech envelope have shown only a group effect (Kong et al., 2014, Vanthornhout et al., 2019). Both such studies aimed to test the impact of attention in general, so they compared the average responses between active and passive listening

conditions. Here, we aimed to measure the presence of responses in each subject to calculate the detection sensitivity with passive listening. Vanthornhout et al. (2019) previously explored the effect of attention on EEG results at a group level for various SNR conditions. They found a significant effect on the part of attention only with low SNR; no difference was indicated for high SNR as was used in the current (noise-free) stimulus. However, from their presented results it was unclear whether the response was significant for all subjects during passive listening. From Figure 1 in Vanthornhout et al. (2019), the correlation value of the decoder in the quiet passive condition ranged from -0.04 to 0.23. This result indicates an absence of response in some participants. This finding is directly in line with our outcome of reduced cortical response detection sensitivity with passive listening.

The average correlation values of the XCOR function found in this work were lower than those of Kong et al. (2014). Kong et al. (2014) investigated the difference in the cross-correlation function between the speech envelope and EEG during active and passive listening in quiet conditions. They found that the average cross-correlation function's peak values were r = -0.03 and r = 0.07 in the active listening condition, and r = -0.03 and r = 0.05 in the passive listening condition. Their passive listening results can be compared to the current study, as both share similar testing settings. The correlation peak values found for the group average were lower, with the maximum positive peak being r = 0.005 for speech and r = 0.01 for speech-modulated noise, while the maximum negative peak was r = -0.01 for speech and r = -0.015 for speech-modulated noise(see Figure 4.7). These differences can be explained by differences in the analysis methods and the sample sizes. Kong et al. (2014) used the averaged EEG responses of ten trials of each stimulus segment to measure the correlation with the speech envelope, while we used single-trial EEG. In ERP, averaging will improve the SNR by reducing unrelated evoked responses (Luck, 2014), which may lead to a higher correlation. In addition, Kong et al. used multiple EEG channels, as compared to the single channel used here. However, it is not clear which channels were included in the analysis. The sample size in Kong et al. (2014) was only eight subjects, compared to 22 subjects in the current study, resulting in higher statistical power. This increase in power allows for more reliable results that can be generalised more effectively.

The role of language processing in neural envelope tracking is debatable. One study has shown that language processing through presenting intelligible speech decreases the accuracy of envelope neural tracking (Zou et al., 2019). Another shows no significant change in envelope tracking when using a speech stimulus played backwards (Howard and Poeppel, 2010). However, a large body of research has shown that the tracking of the speech envelope exhibits a decrease in the envelope reconstruction accuracy with the backward modelling when acoustical degradation affects speech understanding, with increasing SNR (Ding and Simon, 2013, Ding et al., 2014, Etard and Reichenbach, 2019). This suggests that it is due to poorer

understanding, rather than the change in acoustic input. Our findings showed that speech-modulated noise leads to a similar or more robust cortical response than a natural speech stimulus in terms of detection rate and average detection time. This suggests that the detection component of hearing abilities, rather than comprehension or language processing, was tested. One possibility of explaining the difference between the two stimuli is that with modulated noise, there is only a change in envelope, whereas with speech, there are also accompanying changes in spectral cues. The same change in the envelope may lead to different responses, depending on what spectral changes occurred. It is known (e.g., from ASSRs) that frequency modulation can also elicit neural responses.

A significant difference between the two conditions correlation values of the TRF model was seen here (see Figure 4.6 and Figure 4.7). Our results are broadly in line with those of (Zou et al., 2019). Zou et al. (2019) compared the neural tracking of speech envelopes between native Cantonese and foreign speakers listening to Cantonese speech. The prediction accuracy of the TRF model found with high SNR was significantly higher in the foreign group (mean = 0.08) than in the native speaker group (mean = 0.05). Similarly, the prediction accuracy found in the current work was significantly higher for the speech-modulated noise stimulus (mean = 0.044) than for natural speech (mean = 0.038). Attention control could explain the higher correlation values Zou et al. (2019) found compared to the current study.

4.5 Conclusion

Using normal-hearing listeners, the detection of cortical responses to continuous speech and speech-like stimuli during passive listening was less than 100% with a maximum of 15 min of stimulation. Neural tracking of the speech envelope generally showed higher envelope prediction accuracy (correlation coefficients), higher detection rates, and shorter detection times (though not significantly) with speech-modulated noise compared to actual speech. One possible cause of the low detection rate is the passive attention. Therefore, further investigation is needed to confirm the differences in detection rate and time between passive and active listening. Additionally, this study indicates a higher detection rate and lower detection time with the speech-modulated noise stimulus, suggesting its potential as a universal (language-independent) testing stimulus, which could make it a better option for clinical use. The next step will involve exploring the effect of hearing aid processing on the neural tracking of the speech envelope. This will include investigating how hearing aid processing alters the envelope of speech-modulated noise and how this compares to its effects on a natural speech stimulus.

Chapter 5 Quantifying the Effects of Hearing Aid Processing on the Stimulus Envelope

The previous research demonstrated the potential for detecting cortical responses to continuous speech using a single EEG channel, as outlined in the first paper (Chapter 3). The second study (Chapter 4) found that speech-modulated noise exhibits a higher detection rate and shorter detection time than intelligible speech. The following phases of the project were designed to test the feasibility of detecting such responses of aided stimuli. This was done in two studies. The first, which will be discussed in this chapter, was an acoustic measurement of the HA's effect on the stimulus envelope. The second, which will be covered in the next chapter (Chapter 6), was an EEG study using stimuli generated from the current study.

5.1 Introduction

Individuals with normal hearing have a wide dynamic range of approximately 100 dB HL, which ranges between the faintest sounds they can detect and the point at which sounds become uncomfortably loud (Souza, 2016). However, cochlear hearing loss reduces this dynamic range, causing an abnormal perception of loudness known as loudness recruitment (Moore, 2008). To address this issue, most contemporary HAs incorporate compression or automatic gain control (AGC). AGC amplifies soft sounds to compensate for reduced sensitivity in individuals with hearing impairments while reducing the gain for loud sounds to prevent discomfort (Moore, 2008). The compression of the HA is described based on several parameters (Souza, 2002) (see Figure 5.1). The compression knee point or threshold (CT) is the level of intensity where the gain reduction is activated. Before the compression activates, the gain is constant. After activation, with every increase in input, the output increases at a different ratio. The compression ratio (CR) refers to the relationship between the increase in input level and the corresponding increase in

output level. For instance, a compression ratio of 2:1 indicates that for every 2 dB increase in the input signal, the output signal only increases by 1 dB (see Figure 5.1).

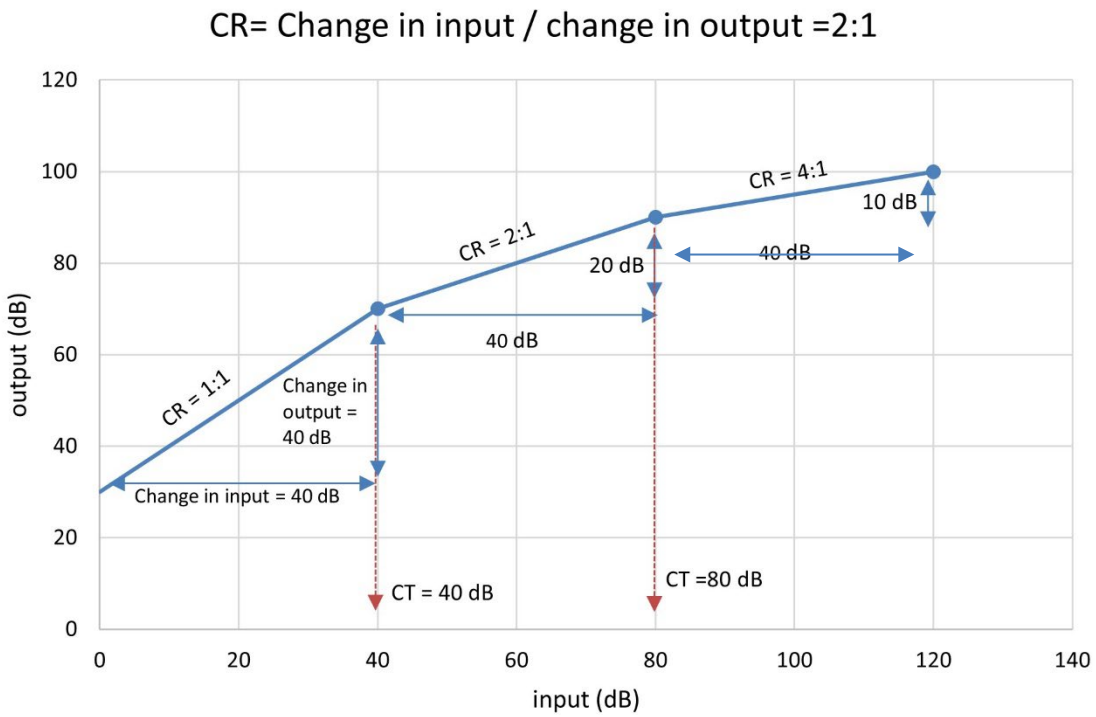

Figure 5.1 A typical hearing aid input/output function. *This figure illustrates the input/output function of the hearing aid in dB, showcasing the compression parameters, specifically the compression threshold and compression ratio. This figure was adapted from Souza (2002).*

Moore (2008) describes two major classes of AGC systems: automatic volume control (AVC) and fast-acting AGC or wide dynamic range compression (WDRC), highlighting the advantages and disadvantages of both systems. The AVC system operates with slow attack and release times of the gain control (500 ms to 20 s), while the fast-acting AGC system has short attack and release times (0.05 ms to 200 ms) to approximate normal loudness perception for HA users (see Figure 5.2). Slow-acting compression, if needed, allows speech to be presented at a comfortable volume, regardless of the initial input level, through a high compression ratio. Generally, slow-acting compression maintains the speech's temporal envelope largely undistorted since the pattern of gains across frequencies changes gradually over time. When using multiple channels, fast-acting compression more effectively compensates for frequency-specific loudness changes than slow-acting compression. AGC quickly improves the audibility of soft sounds immediately following loud ones. Compared to slow-acting compression, fast-acting compression offers better speech intelligibility, especially in the presence of noise (Kowalewski et al., 2018). However, one disadvantage of fast-acting compression is its potentially detrimental effect on the stimulus envelope, reducing loudness contrasts. Overall,

the literature suggests that HA compression systems aim to ensure a better experience for users, though this comes with drawbacks, one of which is the impact on the stimulus envelope, a primary concern in the current project.

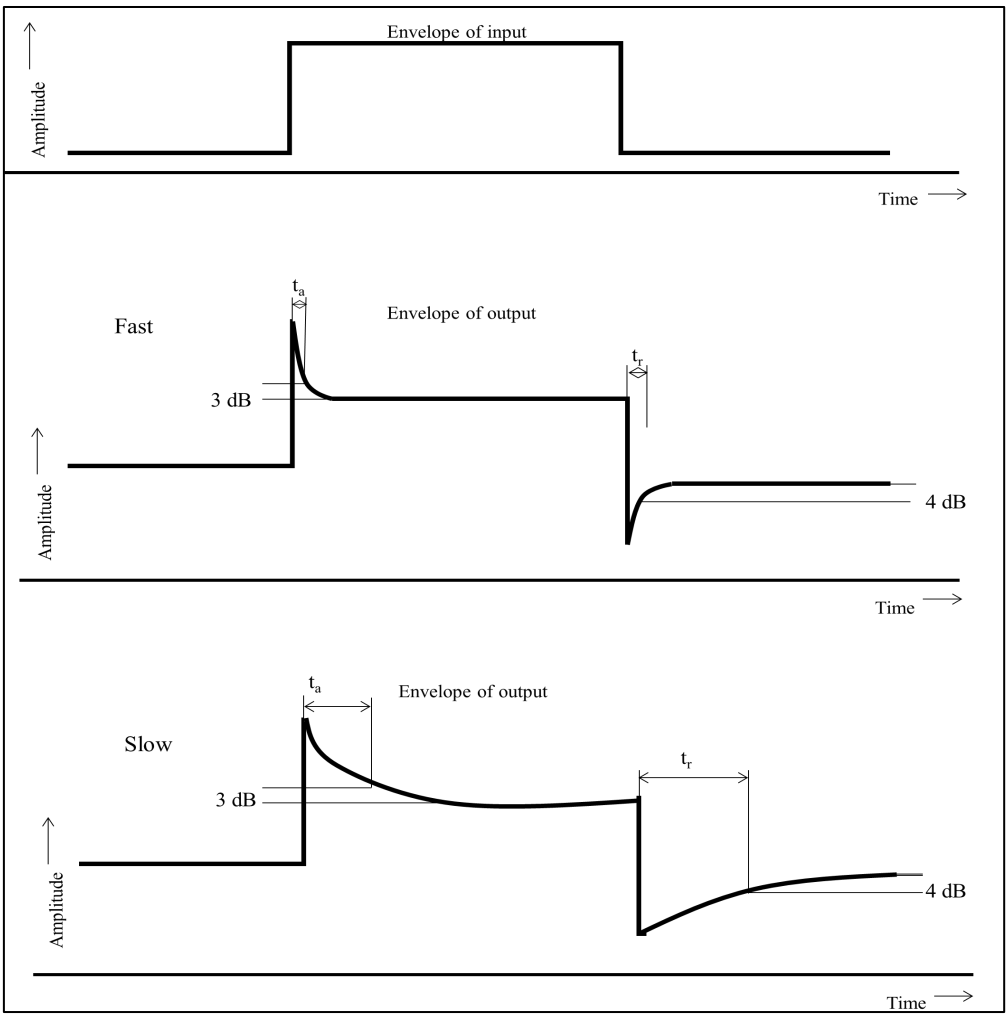

Figure 5.2 The Effect of AGC System on Temporal Response. *The figure illustrates the effect of the fast and slow-acting systems on the input envelope of the signal. The top panel shows the envelope of the input signal. The middle panel shows the fast-acting AGC system and the bottom panel shows the slow-acting AGC system. 'Ta' refers to the attack time, and 'Tr' indicates the release time, both measured in ms. This figure was redrawn from (Moore, 2008).*

The most noticeable temporal change caused by fast-acting WDRC is the distortion of the amplitude envelope, which has received significant attention in the context of HA compression (Souza, 2002). Compression modifies the amplitude envelope of the stimulus, reducing the contrast between speech sounds of high and low intensity (Souza, 2002). Stone and Moore (2007) described various envelope alterations caused by compression. First, they noted a reduction in within-signal modulation correlation, which refers to the correlation between different frequency bands of the signal. Higher correlation helps with perceptually binding the signal components, leading to better speech intelligibility. Second, compression can alter the

fidelity of the envelope, affecting its shape by causing overshooting or undershooting (Verschuure et al., 1996). The more significant the alteration in the shape, the more it can impact speech intelligibility (Stone and Moore, 2007). The higher the CR, the more significant the change will be (Souza, 2002). It is clear that the envelope can be affected by compression in different ways. Since the speech envelope is crucial for speech intelligibility (Shannon et al., 1995) , any distortion might affect understanding (Souza et al., 2012).

As the next step of the current research focuses on measuring envelope neural tracking responses to continuous aided speech, it was essential first to measure the effect of the HAs on the signal envelope. The envelope distortion index (EDI) is widely used to quantify the change caused by HA processing on the stimulus envelope (Jenstad and Souza, 2005, Souza et al., 2012, Geetha and Manjula, 2014, Chinnaraj et al., 2021). EDI measures the difference in the temporal envelope between two signals (Jenstad and Souza, 2005) and was initially developed by Fortune et al. (1994). To determine the amplitude envelope of the aided and unaided signals, the absolute value of the Hilbert transform is first calculated to provide a high-resolution envelope. This is followed by applying a digital low-pass filter with a 30 Hz cutoff (Jenstad and Souza, 2005). The signal is then normalised by 100%. The resulting signals are then scaled by dividing each sample point by the mean amplitude to establish a common reference for comparing the two envelopes. The following formula is used to compute the EDI between the two signals:

Equation 5.1

$$\text{EDI} = \left(\sum_{n=1}^{N} |\text{Env1} - \text{Env2}|\right)/2\text{N},$$

where N represents the number of samples in the signals, Env1 denotes the normalised unaided recorded signal, Env2 corresponds to the normalised aided (processed) signal, and n is the sample index. The resulting EDI values ranged from 0, indicating identical waveforms, to 1, indicating maximally different signals (e.g., perfectly out of phase).

Earlier literature has investigated the effects of various HA settings on the stimulus envelope, including compression release time (Jenstad and Souza, 2005), compression, digital noise reduction (DNR), and directionality (Geetha and Manjula, 2014, Chinnaraj et al., 2021). Jenstad and Souza (2005) examined the impact of three different release times (12 ms, 100 ms, and 800 ms) on vowel-consonant (VC) syllables from the Nonsense Syllable Test (NST) presented at three levels (50 dB, 65 dB, and 80 dB SPL). They used the EDI and the consonant-vowel (CV) ratio as measures of acoustical change. The CV ratio focuses on the phonemic change in the syllable, while the EDI specifically examines the envelope. The results showed EDI values ranging between 0.05 and 0.27 across different conditions. Lower input levels resulted in lower distortion, and no significant impact on release time was indicated. Higher EDI values were

observed at loud input levels (65 dB and 80 dB SPL). This study demonstrated the applicability of the EDI approach in quantifying the impact of specific HA compression features on the stimulus envelope. This observation was based on the EDI calculation's sensitivity to changes in compression settings and input values.

Geetha and Manjula (2014) used EDI to quantify the independent and interaction effects of compression, digital noise reduction (DNR), and directionality on the temporal and spectral features of various speech stimuli. They assessed subjective quality perception as well. The study found no significant effect of the number of HA features activated on the measured distortion, indicating that increased HA digital signal processing does not affect the EDI. However, the study did not provide details about the statistical tests used to evaluate the effect of different algorithms on the EDI. The findings were based on numerical differences between the mean EDI values. The results indicated that a higher presentation level (70 dB SPL) led to increased envelope distortion. This may be due to the higher input intensity resulting in a higher CR, as the HA attempts to maintain the output within a comfortable range for the user, which could explain the increased envelope distortion at higher input levels. Although the study suggests some effect of input level on envelope distortion, the recording setup did not account for the amplification provided by the pinna, as the HA was placed on a tripod rather than on a manikin or a subject's ear. It is important to note that the human auricle contributes to sound amplification around the frequency of 3 kHz (Purves, 2001).

More recently, Chinnaraj et al. (2021) investigated the effect of noise reduction and directionality on EDI and explored the correlation between EDI and speech recognition. They used a 16-channel HA and tested four different conditions: NR on, directionally on, both NR and directionally on, and both NR and directionally off. The study included 20 individuals with mild to moderate sensorineural hearing loss. A KEMAR head and torso simulator (Knowles Electronics Manikin for Acoustic Research) was fitted with a HA programmed for each individual. The results showed significantly higher EDI values for sentences compared to short syllables. Chinnaraj et al. (2021) also found that the effects of NR and directionality varied based on the stimulus level. At the low intensity level (55 dB SPL), there were no significant effects. However, at the medium intensity level (65 dB SPL), the activation of the algorithms increased the EDI. At the high intensity level (85 dB SPL), NR had a greater impact on increasing the EDI compared to directionality. The EDI increased with increasing presentation levels. However, the study did not investigate the statistical differences between presentation levels. Speech recognition scores were high with the activation of the HA algorithms, but no significant correlation was found between EDI and speech recognition. These findings suggest that high-envelope distortion does not necessarily indicate reduced speech perception. Overall, the study shows how EDI can vary with HA conditions. This finding highlights the importance of

quantifying envelope distortion before measuring aided evoked responses, particularly when using continuous speech. Any distortion of the speech envelope may lead to corresponding alterations in the brain's response to the speech. Consequently, using the input envelope to correlate with the EEG response could be compromised, potentially affecting the detection of the response.

In general, previous studies that utilised EDI to quantify the effects of HA processing often relate it to subjective measures such as speech intelligibility, loudness perception, and speech clarity. To the best of our knowledge, the literature does not clearly elucidate how envelope distortion varies with different types of speech and speech-like stimuli. Given the current project's aim to evaluate the feasibility of using cortical responses to continuous speech envelopes for assessing aided speech perception, it was imperative to investigate how HAs alter the envelope of various stimuli. In the preceding study (Chapter 4), it was discovered that speech-modulated noise, which possesses an identical speech envelope to intelligible speech, generated more robust cortical responses, with higher detection rates and shorter detection times than natural speech (see Figure 4.4 and Figure 4.5). As speech-modulated noise could be more suitable for clinical use due to its language independence, it was important to determine whether it behaves in the same manner as normal speech when processed by HAs. Considering the possibility that speech-modulated noise might show different HA effects compared to speech, another language-independent stimulus was included: the ISTS, which was created by using recordings of brief passages of natural spoken human speech from six different languages. The ISTS signal contains core speech characteristics, making it recognisable as speech to humans while remaining unintelligible (Holube et al., 2010). It would be beneficial to the advancement of measuring cortical responses to continuous speech to determine if HAs respond to ISTS in a similar manner to actual speech.

The current study measured envelope distortion with different HA settings to help with the design of the next EEG experiment. Specifically, it included the effects of noise reduction and CR on the EDI for various types of stimuli. This included the two stimuli used in the previous study (see Chapter 4): speech and speech-modulated noise. An NR feature was incorporated as the modulated noise output might be influenced by it. This study assessed how the envelope of the language-independent stimuli (speech-modulated noise and ISTS) is affected by HA processing compared to speech. The study specifically looked at envelope distortion because the stimulus envelope is the feature used to analyse the correlation with EEG responses in the subsequent study (Chapter 6).

5.1.1 Research questions

- Is there a significant difference in EDI values between different stimuli (speech, speech-modulated, and ISTS)?
- Is EDI significantly affected by NR (for each test stimulus)?
- Will increasing the compression ratio affect the EDI?

5.2 Materials and Methods

To achieve the objectives outlined above and provide evidence for selecting suitable stimuli for the next study (Chapter 6), the EDI was calculated for a series of different speech and speech-like stimuli. The stimuli were played through loudspeakers and recorded in the ears of a mannequin. The original and recorded signals were then compared using the EDI index. Details of the methods are provided below.

5.2.1 Recordings Setup

Figure 5.3 illustrates the visual representation of the recording setup. Stimuli were presented through Genelec 8020C loudspeaker positioned at a 0° angle, 1 m away from KEMAR right ear. This loudspeaker was connected to an RME Fireface UC audio interface, which was connected to a laptop running Audacity for presenting the stimuli. A Zwislocki coupler, placed in KEMAR's ear, was used to record the signal. The coupler was linked to a pre-amplifier, which was then connected to an input channel of the RME Fireface. The recorded stimuli were exported as .wav files using Audacity. All equipment was located inside a sound-treated room to mitigate the effect of external noise and to reduce sound reverberation. To ensure consistent recording conditions, the sensitivity of the pre-amplifier was maintained at a constant level of 100 V/Pa throughout the recordings. The gain in the soundcard input for the recorded stimulus was carefully adjusted to prevent peak-clipping in cases of loud input signals. Any gain adjustments were documented to ensure they could be compensated for during stimulus playback in the next study (Chapter 6). The EDI measurement will not be affected, as it will be normalised.

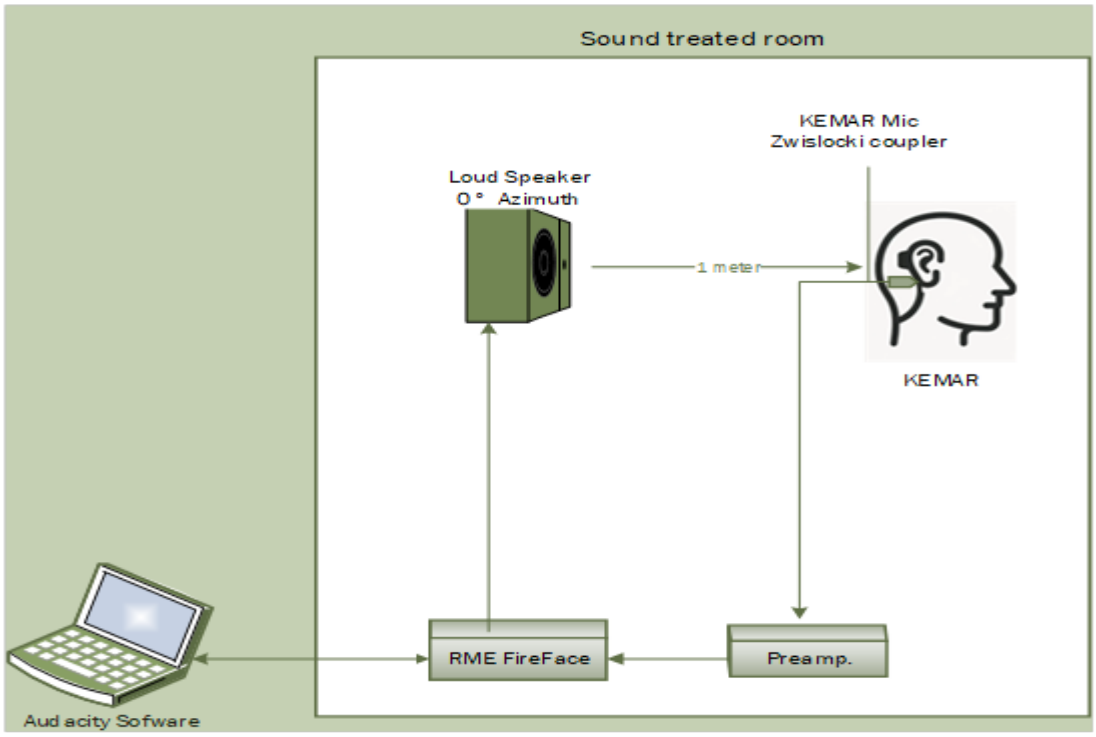

Figure 5.3 Overview of the recording setup. *This figure describes the setup for capturing audio stimuli, detailing the arrangement of a loudspeaker positioned at a 0° angle and 1 m from KEMAR ear, connected to an RME Fireface audio interface and a laptop running Audacity. It shows the pathway from the loudspeaker through the Zwislocki coupler placed in KEMAR's ear for signal recording, connected to a preamplifier and back to the RME Fireface for data capture. All recording equipment except the laptop was located in a sound-treated room.*

This setup was used to conduct both unaided and aided recordings. For the unaided condition, the recording was made directly from KEMAR's ear canal without a HA, aiming to capture the ear canal's effect and thus control for it as a confounding factor when compared to the aided condition. In the aided recordings, KEMAR was fitted with a soft earmold without a vent.

5.2.2 Stimulus

The current study utilised three stimuli: English natural speech, speech-modulated noise, and the ISTS stimuli. Detailed information regarding the speech and speech-modulated noise stimuli can be found in Chapter 4 (see Section 4.2.3). These three stimuli were used to investigate potential differences in their response to HA processing.

Intelligible speech is an ideal option for assessing speech perception, as it closely resembles what individuals typically encounter in daily life. However, for clinical purposes, using a language-independent speech stimulus might be preferable, as it is acoustically similar to

speech but can be used with listeners from different language backgrounds, including pre-lingual infants. Speech-modulated noise and ISTS potentially serve as suitable universal stimuli.

All stimuli lengths were approximately 15 min, allowing for multiple distortion values to be calculated to enable statistical comparisons across different conditions. The ISTS stimulus, initially 1 min, was repeated to have a length of 15 min for this analysis. Each stimulus was presented at sound pressure levels of 50 dB, 70 dB, and 85 dB A-weighted equivalent continuous sound level (LAeq), representing low, moderate, and loud input levels, respectively.

To ensure accurate stimulus levels at the position of the HA microphone when placed on KEMAR's ear, a Sound Level Meter (SLM) was employed. The calibration of the SLM involved a free-field microphone (type 4189 Bruel Bruel & Kjær). It was calibrated using a reference piston at 94 dB SPL. Establishing a reference for acoustic measurements on KEMAR was crucial for ensuring consistent and reliable data collection across various stimuli. To achieve this, a 1 kHz sine wave, selected for its mid-range frequency that is within the human hearing spectrum, was presented at a constant level of 74 dB SPL directly measured at KEMAR's ear.

The stimuli were presented via the loudspeaker positioned at a 0° angle and located 1 m away from KEMAR. The SLM was positioned near KEMAR's ear. The calibration process involved three types of stimuli at three different levels (50 dB, 70 dB, and 85 dB LAeq) to cover a broad range of potential HA responses. Table 5.1 shows the adjustments made to the loudspeaker gain to ensure that the correct sound level was produced at KEMAR's ear for each stimulus (within a 1 dB tolerance). The 1 dB was tolerated to ensure that the loudspeaker gain remains consistent across different presentation levels, which helps maintain uniformity and reliability in the measurements. To verify the reference accuracy, with the loudspeaker adjusted to -14 dB, the 1 kHz sine wave consistently produced a measured sound level of 74 dB SPL, while other stimuli resulted in sound levels around 70 to 71 dB LAeq. The difference in levels is expected due to the acoustic differences between speech and noise-like stimuli compared to a sine wave (Hansen, 2001). However, as the various stimuli fall within a consistent and expected range relative to the reference signal, the calibration can be considered effective.

Table 5.1 Stimuli levels measured using Sound Level Meter (SLM) *Placed Next to the Ear of KEMAR. This table provides the measured levels of various stimuli used in the study, captured with a sound level meter positioned next to the ear of the KEMAR.*

Loudspeaker Gain	Stimulus Type	Stimulus level at KEMAR
-14 dB	1 kHz sinewave	74 dB SPL
-14 dB	Speech	70.5 dB LAeq
-14 dB	Modulated noise	71.2 dB LAeq
-14 dB	ISTS	71.2 dB LAeq
-34 dB	Speech	50.5 dB LAeq
-34 dB	Modulated noise	51.4 dB LAeq
-34 dB	ISTS	51.5 dB LAeq
+1 dB	Speech	85.2 dB LAeq
+1 dB	Noise	85.6 dB LAeq
+1 dB	ISTS	86 dB LAeq

5.2.3 Hearing aid and testing condition

The HA used in this study was Oticon Engage behind-the-ear (BTE) HA. The chosen HA was programmed using Genie software (the official software from Oticon to programme) for moderate flat sensorineural hearing loss, with thresholds ranging from 50 dB to 60 dB HL. During the programming process in Genie, the default settings of the General DSL v5A-Paediatric prescription were initially utilised. Following that, three different gain adjustments were made to have three different compression settings (low CR, high CR and linear gain). Table 5.2 provides an overview of the three different compression settings and the corresponding HA gains. For the low compression programme, the HA gain was adjusted based on measurements obtained from testing the HA in the test box – the resulting CRs were 1.8:1 and 2.1:1. These gain adjustments were made to provide a ~20 dB gain for soft sounds (50 dB), a ~10 dB gain for moderate sounds (70 dB), and a ~5 dB gain for loud sounds (85 dB) as measured in the test box. The gain was further modified for high CR CRs = 2:1 and 3.1:1. The modification was based on having higher compression but ensuring a similar overall gain compared to low compression conditions (+/-2 dB) through test box measurement. For the linear condition, the gain was adjusted to be 20 dB for all testing conditions to ensure a zero compression or constant gain.

Table 5.2 Gain and compression ratios (CRs) for each testing condition. *This table provides a detailed summary of the gain settings and compression ratios (CRs) applied under each testing condition for the hearing aid. Table A presents the low compression settings, Table B shows the high compression settings, and Table C outlines the linear compression settings. The first compression threshold (CT) is at 50 dB SPL, the second CT is at 65 dB SPL, and the third CT is at 80 dB SPL.*

Table A															
Frequency (Hz)	125	250	500	625	750	1K	1.25	1.5	1.75	2K	3K	4K	5K	6K	Average
Loud	5	10	12	12	12	11	12	10	10	9	7	5	7	8	9.3
Moderate	11	17	19	19	19	18	18	17	16	5	14	12	13	12	15.0
Soft	24	30	27	27	30	32	32	31	31	30	27	25	27	27	28.6
CR	1.9	1.9	1.9	1.9	1.9	1.8	1.7	1.9	1.7	1.7	1.9	1.9	1.7	1.4	1.8
CR	2.3	2.9	2.0	1.9	2.1	2.4	2.3	2.2	2.2	2.1	1.9	1.8	1.7	1.9	2.1
Table B															
Frequency (Hz)	125	250	500	625	750	1K	1.25	1.5	1.75	2K	3K	4K	5K	6K	Average
Loud	5	9	8	8	8	9	9	9	8	7	5	5	5	6	7.2
Moderate	10	15	15	16	16	16	15	15	14	15	14	13	14	14	14.4
Soft	24	30	27	28	31	34	33	33	34	34	31	31	31	31	30.9
CR	1.7	1.7	1.9	2.1	2.1	1.9	1.7	1.7	1.7	2.1	2.5	2.1	2.5	2.1	2.0
CR	2.9	4.0	4.0	3.4	3.5	3.3	3.6	3.3	3.5	2.9	2.5	2.6	2.1	2.2	3.1
Table C															
Frequency (Hz)	125	250	500	625	750	1K	1.25	1.5	1.75	2K	3K	4K	5K	6K	Average
Loud	15	21	20	20	20	20	20	20	20	20	19	20	19	20	19.6
Moderate	15	21	20	20	20	20	20	20	20	20	19	20	20	20	19.6
Soft	15	21	20	21	21	21	20	21	21	20	20	20	20	22	20.2
CR	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.0
CR	1.0	1.0	1.0	1.1	1.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Different HA programs were employed for each of the testing conditions. The primary tested features included noise reduction with the option to turn it on (NRON) or off (NROFF), and the CR with choices of low or high settings. The conditions tested were as follows:

1. Low CR + NRON
2. Low CR + NROFF
3. High CR + NROFF
4. Linear CR + NROFF

For investigating the NR effect, only the low CR was utilised. The chosen combination of HA programs, compression settings, and NR options enables the assessment of the impact of different signal processing features on the signal envelope and subsequent analysis of their effect on measuring cortical responses. To evaluate the impact of HA processing, recordings were made for each stimulus and level under unaided conditions.

5.2.4 Acoustic analysis

The method chosen to quantify the envelope distortion involved measuring EDI values, computed using MATLAB code. As described earlier, EDI is a measurement used to quantify the difference between processed and unprocessed stimuli. In this study, we have implemented a similar method described by Jenstad and Souza (2005). The reference signal used to calculate the EDI was the recorded stimulus without the hearing aid (unaided). The two envelopes were first normalised by dividing each one by the mean amplitude of the signal, followed by alignment to account for the delay caused by the HA processing. Alignment was achieved by finding the delay that produced the maximum cross-correlation value between the two signals. After alignment, the EDI was calculated according to Equation 5.1 as described in Section 5.1

The computed EDI was applied to segments of the recorded signals to compare the EDI values among different conditions. For each condition, data were analysed in 24 segments of 30 s duration.

5.3 Results

5.3.1 Visualising envelope distortion:

Before showing the compression of EDI values between different stimulus types, the aided and unaided envelopes were visually inspected. Figure 5.4 shows the normalised envelope of the aided (red) and unaided (black) signals for the three stimulus types used in the study.

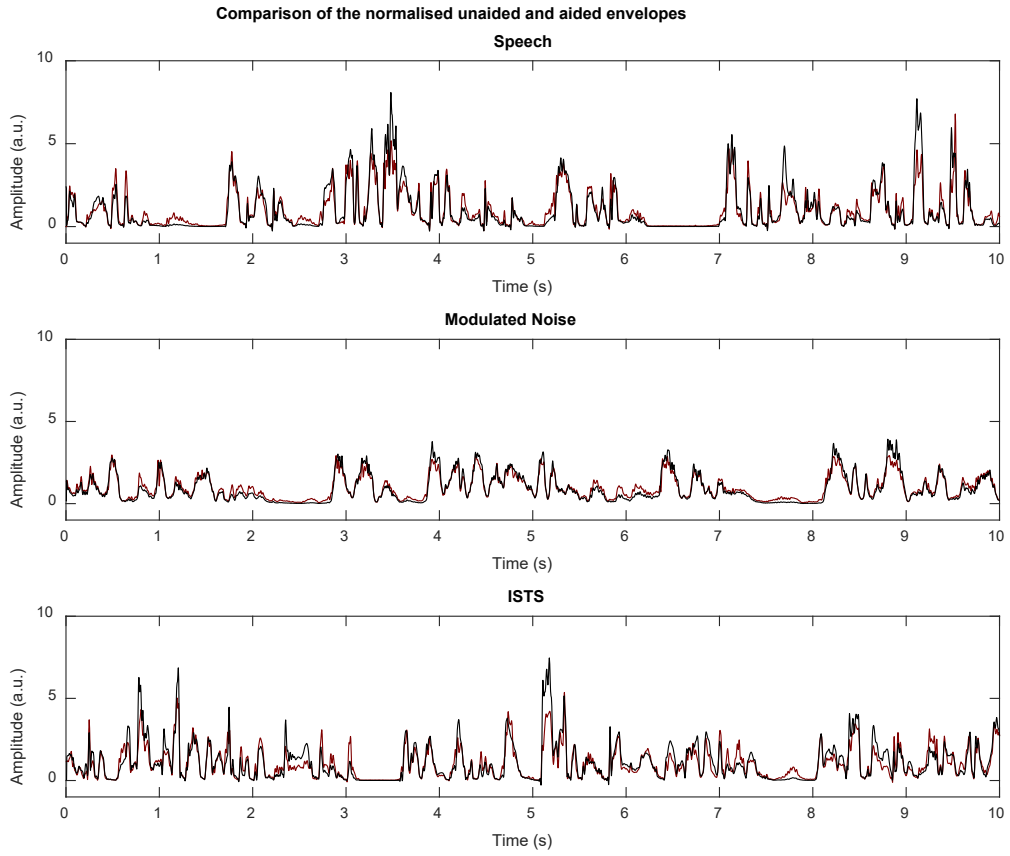

Figure 5.4 Normalised unaided and aided envelope signals for different stimuli. *The figure shows a 10 s envelope of the unaided signal in black and the aided signal in red. The top panel shows the speech stimulus, the middle panel shows the modulated noise, and the bottom panel shows the ISTS stimulus. This was the result of the 85 dB LAeq level condition without noise reduction (NROFF).*

When examining the modulated noise stimulus (middle panel), although the differences are small, there is a minimal compression effect at the peaks with higher levels, where the aided envelope is slightly lower in amplitude (see peaks between 8 s and 9 s). In contrast, the speech stimulus (top panel) shows a more variable compression effect. This variability is evident for speech at peaks between 9 s and 10 s, and for ISTS, between 5 s and 6 s. Additionally, for the speech and ISTS unaided stimuli, there are some negative values in the aided envelope compared to the modulated noise (e.g., speech between 3 s and 4 s and ISTS ~5 s). This overshooting might result from the filter applied during the envelope extraction step (see Section 5.2.4). In contrast, the modulated noise does not exhibit any overshooting.

5.3.2 Effect of NR and Stimuli Type on EDI

Before expanding on the EDI results, it is important to highlight that most previous studies on EDI have focused on short speech syllables. For example, Jenstad and Souza (2005) used the

Nonsense Syllable Test, which consisted of 25 items. To the best of my knowledge, the use of long-running speech as a stimulus in this type of analysis has not been tested before. Consequently, no prior studies are available for direct comparison. The primary aim of this analysis was to assess how these stimuli function in the context of everyday conversation, where a 30-second segment—containing approximately 6–7 sentences—can effectively represent continuous speech, the intended focus for cortical response testing.

Additionally, the speech and speech-modulated noise stimuli used in Chapter 4 were 15 minutes long and structured into four blocks, each lasting 3 to 4 minutes. To ensure sufficient data for analysis while avoiding the risk of exhausting stimuli at the end of each block, 30-second segments were selected from each block, resulting in 24 segments. However, the 15-minute duration of the ISTS stimuli introduces a potential limitation: ISTS repeats every 1 minute, meaning the selected segments are not fully independent samples. A more detailed discussion of this limitation and its implications is provided in the Discussion section.

To answer the first two questions about the effect of NR and stimulus type on the measured EDI, a comparison was made between low CR conditions of either NRON or NROFF with different stimulus types. Shapiro-Wilk tests of normality were conducted for all test conditions, indicating that the data were generally normally distributed, supporting the use of parametric tests. Figure 5.5 shows the mean values of EDI across 24 recording segments in each condition. Here, the analysis examined the impact of NR (NRON and NROFF) and stimuli type (speech, speech-modulated noise and ISTS) on the dependent variable EDI, with an additional consideration of different stimulus levels (50 dB, 70 dB and 80 dB LAeq). This was achieved using a three-way repeated measures ANOVA.

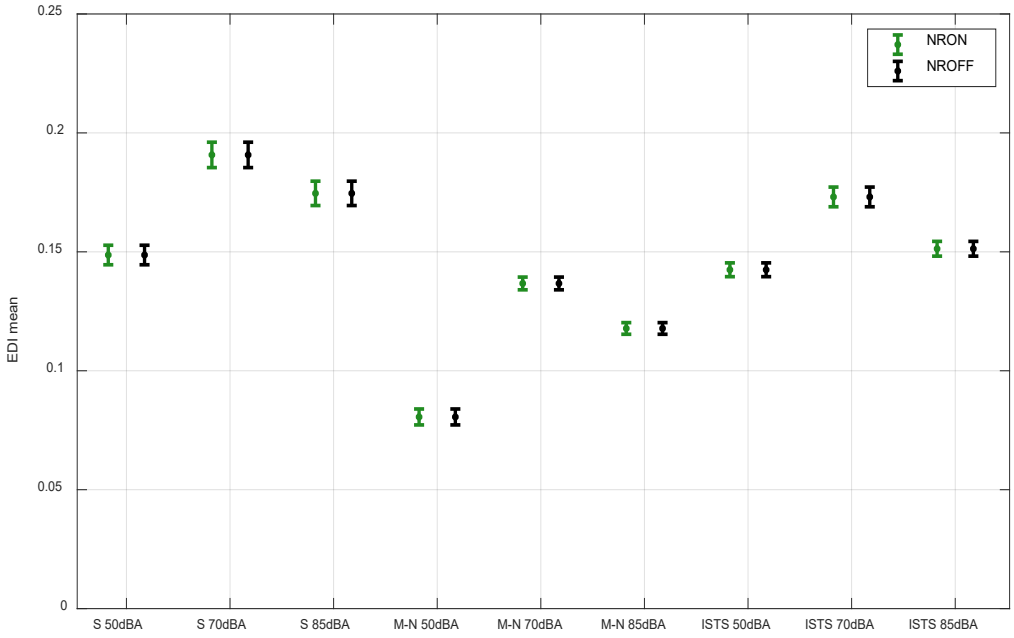

Figure 5.5 Envelope distortion index (EDI) with noise reduction on (NRON) and off (NROFF) for three different stimuli across three stimulation levels*. The figure shows the error bars (mean ± 1 SE of the mean) of the EDI values with and without noise reduction activated. "S" refers to speech, "M-N" refers to speech-modulated noise and the ISTS is the international speech testing signal. Each stimulus was measured at three input levels: 50 dB, 70 dB, and 85 dB LAeq. The compression remained consistent at a low CR.*

The results of the three-way ANOVA on the effect of NR, stimuli types and stimuli levels indicate that, with a Greenhouse-Geisser correction, the main effect of NR on the EDI was not statistically significant ($F(1, 23) = 1.690$, $p > 0.05$). However, the main effect of stimulus type was found to be statistically significant ($F(1.42, 32.66) = 74.98$, $p < 0.0001$). A Bonferroni post hoc test showed that the modulated noise stimulus had a statistically significantly lower EDI mean (0.114 +/- 0.003) than speech (0.17 +/- 0.005) and ISTS (0.155 +/- 0.003) ($p < 0.0001$). No significant difference was indicated between speech and ISTS. These findings were evident for all stimulation levels, except at 85 dB LAeq ISTS, which showed significantly lower EDI (0.15 +/- 0.003 than speech (0.17 +/- 0.005) ($p < 0.01$).

A significant interaction was found between NR and stimulus type ($F(2, 46) = 27.5$, $p > 0.05$). A Bonferroni post hoc test showed that the difference between NR conditions (NRON and NROFF) in EDI was significant with all stimulus types (see Figure 5.6). The direction of NR's effect depends on the stimulus type used. With speech and ISTS, the EDI was significantly lower with NRON (0.168 +/- 0.005) and (0.153 +/- 0.003) than NROFF (0.171 +/- 0.005) and (0.156 +/- 0.003), respectively ($p < 0.001$). The effect was the opposite with modulated noise, as NRON has a higher EDI (0.116 +/- 0.004) than NROFF (0.112) ($p < 0.001$).

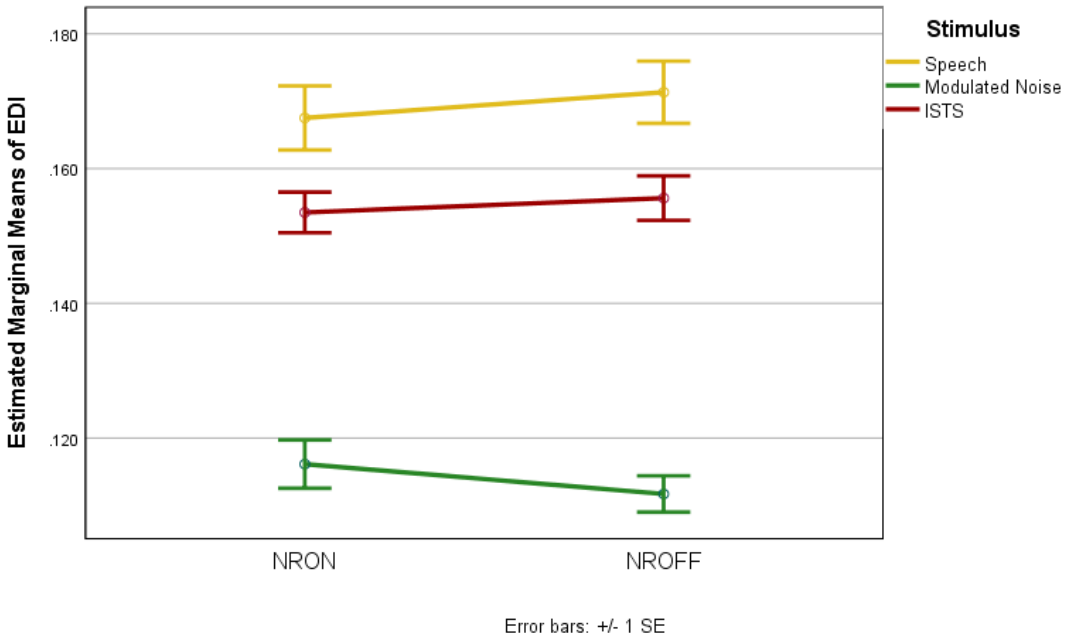

Figure 5.6 Envelope distortion index (EDI) estimated marginal means for the interaction effect between noise reduction (NR) and stimulus type*. The x-axis represents the NR setting, while the bars on the y-axis show the estimated marginal means of EDI, averaged across the three stimulus levels: 50 dB, 70 dB, and 85 dB LAeq. The line colours correspond to different stimuli: the yellow line represents the speech stimulus, the green line represents the speech-modulated noise stimulus, and the red line represents the ISTS stimulus.*

5.3.3 Effect of Stimulus Input Level on EDI

Additionally, the effect of stimulus level on envelope distortion was also investigated. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi 2(2) = 1.70$, $p = 0.43$. The main effect of the stimulus level on EDI was statistically significant ($F(2, 46) = 989.44$, $p < 0.0001$). A Bonferroni post hoc test showed that 50 dB LAeq level showed a significantly lower EDI (0.125 +/- 0.003) than 70 dB LAeq and 85 dB LAeq ($p < 0.001$). In addition, 70 dB LAeq A produced significantly higher EDI (0.166 +/- 0.003) than 80 dB LAeq (0.147 +/- 0.003), ($p < 0.001$).

With NR and level interaction, the effect on EDI was significant ($F(2, 46) = 5.03$, $p > 0.05$). The Bonferroni post hoc test showed that the difference between NR conditions was only significant at 85 dB LAeq level as NRON EDI mean (0.148) was higher than NROFF EDI (0.146) ($p < 0.01$) (see Figure 5.7). It is important to note that while the interaction effect was statistically significant, higher EDI with NRON at 85 dB LAeq level, its magnitude was relatively low (0.002).

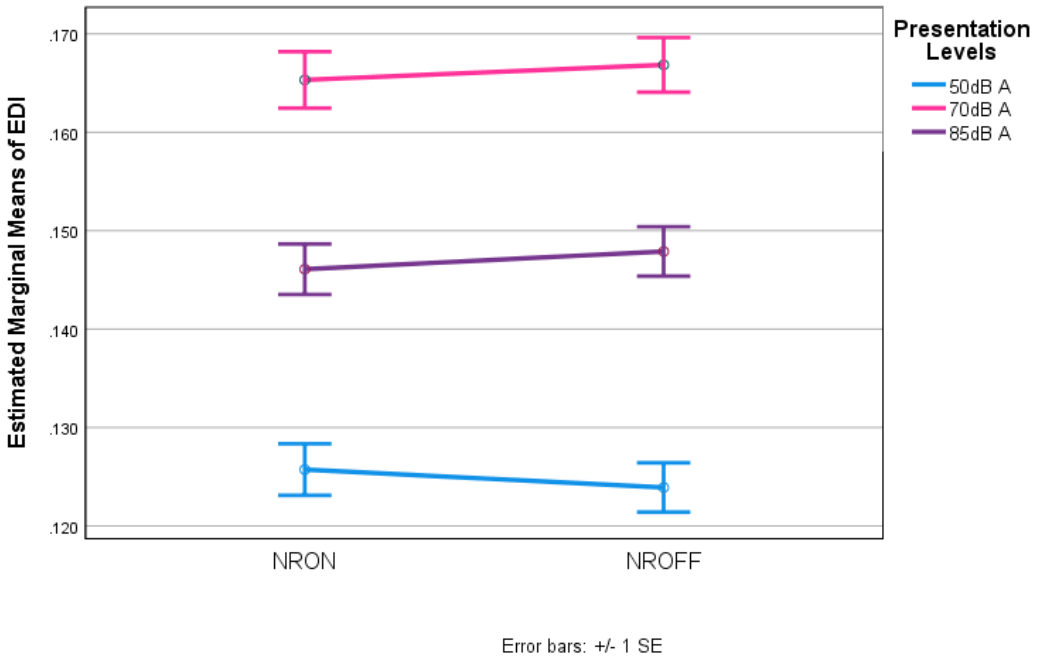

Figure 5.7 Envelope distortion index (EDI) estimated marginal means for the interaction effect between noise reduction (NR) and stimulus level*. The x-axis represents the NR setting, while the bars on the y-axis represent the estimated marginal means of EDI averaged across the three stimuli types. The line colours correspond to different input levels: the blue line represents the 50 dB LAeq, the pink line represents the 70 dB LAeq, and the purple line represents the 85 dB LAeq.*

Since the results showed that 70 dB LAeq had a higher EDI than 85 dB LAeq, visual inspection was conducted, as more compression and distortion were predicted at 85 dB LAeq. Figure 5.8 shows the normalised envelope comparing the two input levels with the speech envelope. In the figure, larger compression effects can be observed at 70 dB LAeq A, indicated by the circles around the peaks.

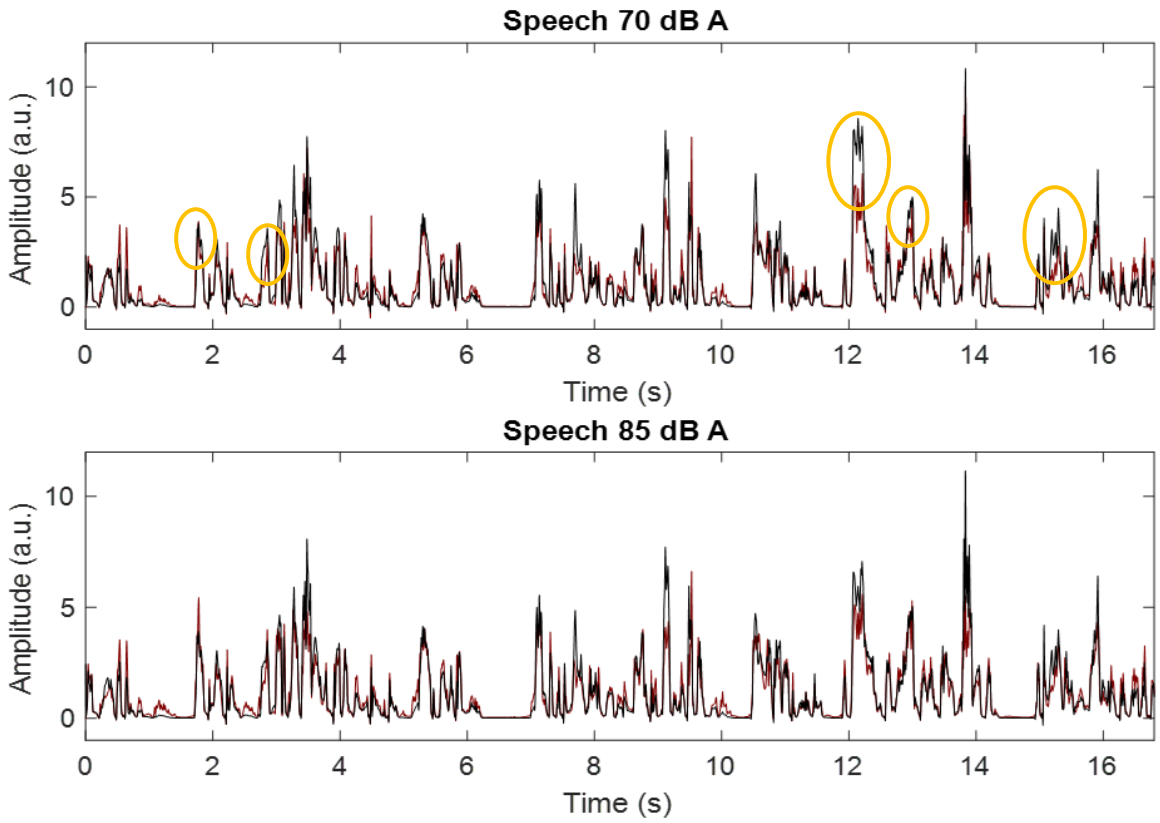

Figure 5.8 Normalised unaided and aided envelope signal for different input levels. *The figure shows a 16 s envelope of the unaided signal in black and the aided signal in red for the ISTS stimulus. The y-axis represents the amplitude of the signal, while the x-axis represents time in seconds. The top panel shows the ISTS at a 70 dB LAeq input level, and the bottom panel shows the ISTS at an 85 dB LAeq input level.*

5.3.4 Effect of compression ratio on envelope distortion index

Figure 5.9 shows the mean values of the EDI for the three compression settings. To address the third research question about the effect of different CR on the envelope distortion, the EDI levels for different stimuli and input levels were compared across three compression settings: low, high, and linear (see Table 5.2), where The CR in linear compression was set to 1. A repeated measures ANOVA with Sphericity assumption was performed, revealing a statistically significant main effect of compression on EDI ($F(2, 46) = 628.28$, $p < .0001$). The Bonferroni post hoc test indicated that the mean EDI with high CR (0.177 +/- 0.002) was significantly higher than that with low CR (0.146 +/- 0.002) and linear CR (0.135 +/- 0.002), with a significance level of $p < 0.001$. The linear CR showed significantly lower EDI than low CR, $p < 0.001$.

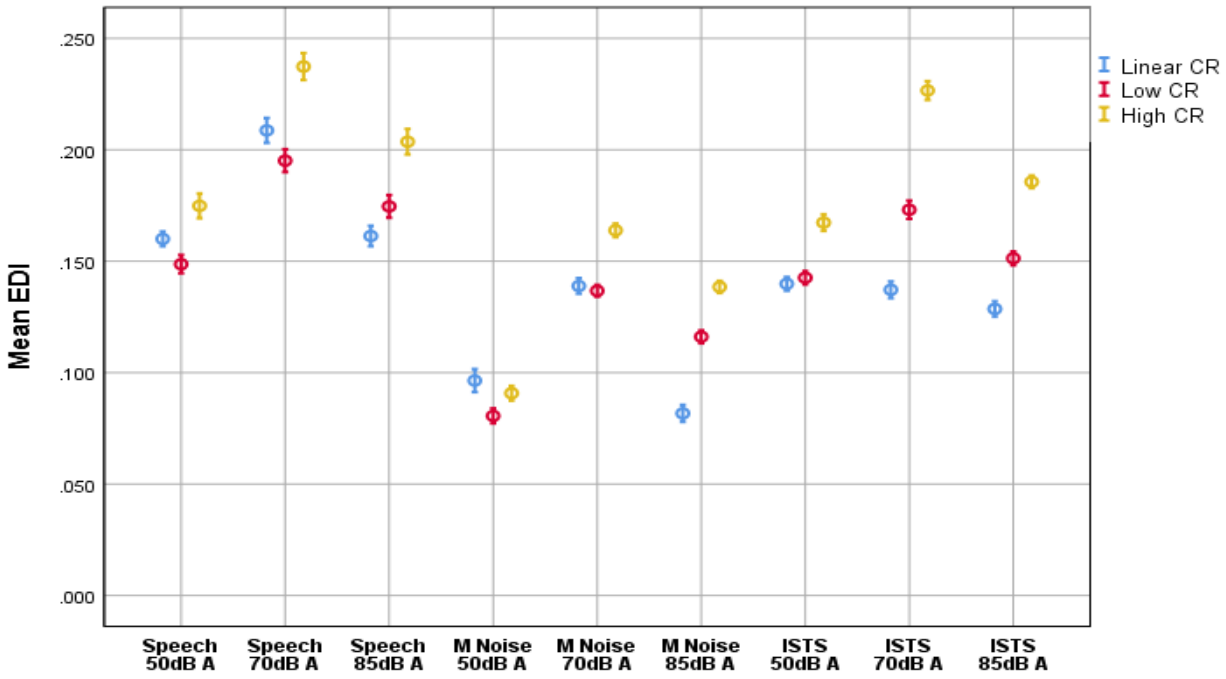

Figure 5.9 Envelope distortion index (EDI) at different compression ratios (CR). *This figure displays mean EDI and error bars (mean ± SE) for three different CRs: Low, High, and Linear. For each CR, three stimuli (Speech, Speech-modulated noise, and ISTS) and three stimuli levels (50 dB, 70 dB, and 85 dB LAeq) were used. The bar colours correspond to the different CRs: blue bars represent Linear CR, red bars represent Low CR, and yellow bars represent High CR.*

Furthermore, the interaction between compression and input level was found to be significant with Greenhouse-Geisser correction ($F(2.9, 66.9) = 233.08$, $p < 0.0001$) (see Figure 5.10). Specifically, the Bonferroni post hoc test indicated that the high CR demonstrated significantly higher EDI values compared to the other compression settings at all input levels ($p < 0.001$). Additionally, the linear CR exhibited significantly lower EDI values at 70 dB LAeq and 85 dB LAeq compared to the low CR ($p < 0.001$), indicating an interaction effect between compression settings and input levels.

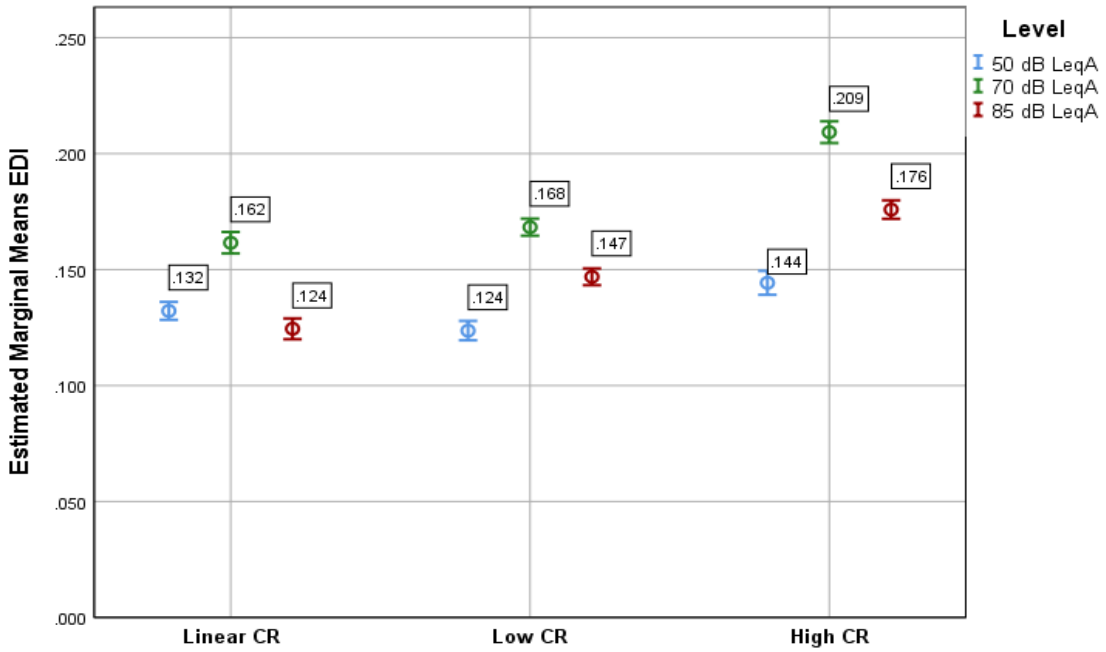

Figure 5.10 Envelope distortion dndex (EDI) means for compression ratio (CR) and stimulus Level. *The figure presents the error bars (mean ± 1 SE) for the estimated marginal EDI means, showing the interaction effect between the three CRs (Low, High, and Linear) and input levels (50 dB, 70 dB, and 85 dB LAeq). The x-axis represents CR, while the bars on the y-axis represent the estimated marginal means of EDI average of three stimuli types. The bar colours correspond to different input levels: blue bars represent 50 dB LAeq, green bars represent 70 dB LAeq, and red bars represent 85 dB LAeq.*

The analysis of the interaction between compression and stimulus type, with Greenhouse-Geisser correction, revealed a significant effect on the EDI ($F(2.3, 52.8) = 62.15$, $p < 0.0001$) (see Figure 5.11). Specifically, as expected, the high CR exhibited higher EDI values compared to the other compression settings for all stimuli ($p < 0.001$). On the other hand, the linear CR showed lower EDI values for the modulated noise and ISTS stimuli compared to the low CR ($p < 0.001$), indicating a significant interaction effect between compression setting and stimulus type.

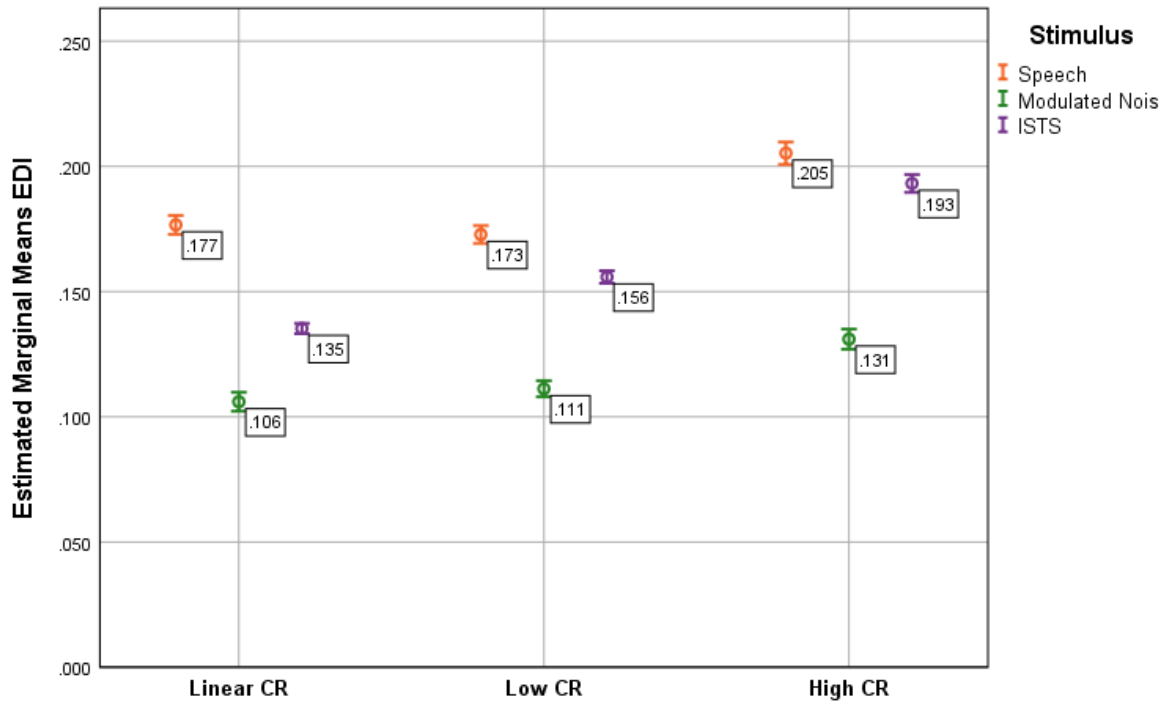

Figure 5.11 Envelope Distortion Index (EDI) Means for compression ratio (CR) and stimulus type. *The figure presents the error bars (mean +/- 1 SE) for the estimated marginal EDI, showing the interaction effect between the three CRs (Low, High, and Linear) and stimuli type (Speech, Speech modulated noise, and ISTS). The x-axis represents CR, while the bars on the y-axis represent the estimated marginal means of EDI average of three stimuli levels. The bar colours correspond to different stimuli: the orange bars represent speech, the green bars represent speech-modulated noise, and the purple bars represent the ISTS.*

5.3.5 Recording quality

To ensure accurate measurements, repetitions of unaided recordings were conducted during the pilot study to verify that the noise floor in the room did not introduce any distortion. Subsequently, the EDI was computed using a single stimulus segment. The mean EDI values obtained from the repetitions consistently remained low across the measurements (0.0049, 0.0048, 0.0056, 0.0055, 0.0045, 0.0047), indicating minimal influence of room noise.

5.4 Discussion

This study investigated the effect of HA processing on the signal envelope by examining different stimulus types and hearing aid settings by measuring the amount of envelope distortion (using EDI). The results showed that speech-modulated noise exhibited significantly lower distortion compared to speech and ISTS signals. The compression ratio had a significant main effect on EDI, with higher compression leading to a significant increase in EDI. Additionally, a moderate input level produced significantly higher envelope distortion compared to low and high levels.

The effect of noise reduction on EDI was significant only when considering the interaction with stimulus type or stimulus level.

The impact of NR varied depending on the specific stimulus used. Modulated noise showed higher EDI values with NRON, while speech and ISTS had lower EDI values with NRON. This difference may be attributed to how the NR algorithm operates. A HA with adaptive NR aims to provide less amplification for frequency ranges where the noise is intense (National Institute for Health and Care Excellence (NICE), 2018). It is clear from this result that noise reduction affects speech-modulated noise differently than natural speech. These findings emphasise the importance of considering NR settings when choosing the stimulus type to evaluate the impact on envelope distortion.

As mentioned earlier, the EDI measured from speech-modulated noise was significantly lower than speech and ISTS, with and without NR (see Figure 5.6). Additionally, modulated noise consistently showed lower distortion across different compression ratios (see Figure 5.9). Visual inspection of the aided and unaided envelopes suggested that modulated noise had a less compression effect than speech and ISTS (see Figure 5.4). One possible explanation is that speech exhibits more variable spectral information. The HA used in this study was a multichannel model, where compression varies across frequency bands (Dillon, 2012). As a result, spectral changes can alter the envelope. For instance, high gain in high-frequency bands enhances the envelope due to fricatives, while lower gain in low-frequency bands reduces the envelope for vowels. Even pure amplification can lead to these changes in the speech envelope. On the other hand, modulated noise is smoother, which may be related to how it is generated. The speech envelope used to create the modulated noise was smoothed by a low-pass filter (see Section 4.2.3), as well as the filter applied to extract the modulated noise envelope when calculating the EDI.

When using natural speech to measure envelope distortion, it becomes challenging to differentiate between the HA's impact on temporal fluctuations and other potential side effects, such as spectral distortion (Souza, 2002). This may also account for the higher envelope distortion values observed with speech stimuli. When comparing ISTS and speech stimuli, ISTS had lower envelope distortion, but the difference was not statistically significant. Since the distortion of ISTS was not significantly different from that of speech and exhibited similar behaviour with the noise reduction effect, using ISTS might be more appropriate as it more closely resembles actual speech.

Envelope distortion was significantly different between different compression settings; the high CR resulted in significantly higher EDI values than low and linear compression (see Figure 5.9). This observation aligns with previous findings reported by Souza et al. (2012). In their study,

Souza et al. (2012) measured the EDI for syllables using four different CRs ranging from linear to 10:1. They found that the mean EDI increased as the compression ratio increased. Here, three different compression settings were compared: linear, which has a CR of 1:1; low compression with two CRs (1.8:1 and 2.1:1); high compression with two CRs (2:1 and 3.1:1). Similar to Souza et al. (2012), EDI progressively increased from a linear to a higher value, although only a limited range was tested here. Compression affects speech by reducing the amplitude variation, lowering the level of high-intensity phonemes relative to low-intensity phonemes, and resulting in amplitude smoothing (Souza, 2002). The phoneme of amplitude smoothing was explained by Dillon (2012), as a reduction of inter-syllabic and inter-phonemic intensity difference, which results in a narrow dynamic range between the envelope of soft and loud signal. This effect is more pronounced with higher compression ratios, as shown in the current results (see Figure 5.9), particularly with speech and ISTS stimuli. The findings in this study regarding the different effects of compression on envelope distortion will inform the design of future EEG experiments, particularly in investigating the correlation between envelope distortion values and cortical response correlation values.

The EDI showed higher values with a 70 dB LAeq level of presentation compared to 85 dB LAeq (see Figure 5.6). This result contrasts with Chinnaraj et al. (2021), who found higher EDI values at 85 dB SPL compared to 65 dB SPL. A possible explanation could be attributed to the functioning of the Oticon HA compression system, which operates in two primary ways. One approach focuses on the long-term level of the signal, while the other detects rapid sound level variations that exceed typical speech signal shifts (Oticon, 2023). This system was explained by Moore (2008) as the following: “suppose there is a sudden increase in sound level so that the momentary output level rises by more than 8 dB above the level of the ongoing ‘running’ signal level determined by the slow system”. In that case, the fast-acting control signal rapidly reduces the gain, thus avoiding uncomfortable loudness (Moore, 2008). In this context, the HA may activate the fast-acting compression more frequently for moderate-level signals (70 dB LAeq) as the stimulus peaks level can be higher relative to the running level compared to loud signals (85 dB LAeq) (see Figure 5.8). In the louder condition, the overall dynamic range is more compressed, leading to fewer peaks exceeding the running level. As a result, the fast-acting compression is activated less frequently. This observation provides insight into how the HA’s compression system adapts to input level changes, potentially contributing to the differences observed in the EDI values between the two presentation levels.

The repeated measures ANOVA used in this analysis has a key limitation: the ISTS segments are not entirely independent samples, as the stimulus repeats every minute. While this non-independence is a potential concern, some variability is still present in the results. When generating this stimuli, each 1-minute segment was joined to the previous one to ensure

continuity and avoid gaps. Upon closer inspection, there is a slight overlap of 1–2 seconds between the 1-minute segments, which could serve as one source of variability in the envelope distortion. Other potential sources of variability include background noise or internal electronic noise. These were quantified in the study through repetitions of unaided recordings. Specifically, two unaided recordings were made with the hearing aid present but turned off, and the EDI was calculated. The EDI values ranged between 0.001 and 0.01. Although these values are low, they may have contributed to some degree of variability in the measurements.

One possible approach could be to analyse the entire 1-minute ISTS stimulus as a single unit and compare its EDI to that of the full-length speech and speech-modulated noise conditions. This would provide an estimate of the population result for each stimulus. However, in the next phase of the project, the plan is to present the stimuli for 15 minutes with 1-minute increments. Calculating the EDI as a whole does not account for potential variability over time, particularly in speech and modulated noise, where gradual changes may occur. Similarly, for ISTS, using only the first minute does not consider potential differences in EDI values across subsequent minutes. This introduces a limitation to this method.

5.5 Conclusion

In conclusion, this chapter helps to understand how HA speech processing affects the amplitude envelope. It is clear that HA settings significantly impact the envelope, with compression ratio and stimulus level having notable effects on the EDI. Noise reduction showed minimal influence. Complex interactions between stimulus type, level, and compression effects were observed. These factors should be carefully considered when selecting HA settings for the next experiment, which will involve measuring EEG responses to continuously aided stimuli. The impact of envelope distortion on cortical responses will be assessed by comparing different distortion levels on detection sensitivity, detection time, and neural tracking correlations. Parameters will be selected based on these findings as the following:

1- Although it had a minor effect, NR will be deactivated to control it as a confounding variable, as the result suggests that it has inconstant impacts on the level of envelope distortion measured depending on the stimuli.
2- Given that the distortion was influenced by compression and input level due to variation in EDI across conditions, conditions with different EDI values will be employed to evaluate whether the extent of envelope distortion affects the measured cortical responses.
3- The ISTS will be used as the primary stimulus. This decision is based on the observation that ISTS exhibited similar behaviour to speech in terms of the effects produced by HA

processing, compared to speech-modulated noise. Additionally, ISTS is language-independent, allowing for universal use and avoiding comprehension as a confounding factor. In contrast, speech-modulated noise showed significantly lower distortion values than intelligible speech.

Chapter 6 Investigating the Effect of Hearing Aid Envelope Distortion on Cortical Responses to Continuous Speech

After quantifying the effect of HA processing on continuous speech stimuli (Chapter 5), the next step will be to investigate how this is reflected in the cortical response. Therefore, the experimental design described in this chapter was developed based on the findings from the previous study (Chapter 5) and will investigate how the brain processes the HA-modified stimuli generated in that study.

6.1 Introduction

While several studies have explored neural tracking of speech envelopes in individuals with hearing loss — primarily adults (Petersen et al., 2017, Vanheusden et al., 2020, Decruy et al., 2020) and recently in children (Van Hirtum et al., 2023) — the specific relationship between envelope distortion and cortical responses has not been previously investigated. Petersen et al. (2017) conducted one of the pioneering studies on neural responses to aided continuous speech in individuals with hearing loss. The research detected cortical responses from elderly participants with various degrees of hearing loss under different SNR conditions using a cross-correlation approach. The results revealed a significant cortical response to both attended and ignored stimuli at a group level. While attended speech showed significantly stronger neural tracking with a higher correlation value than ignored speech, the disparity diminished in participants with severe hearing loss. The observation was explained by a deficiency in the ability to segregate between competing talkers, resulting in a reduction of the inhibition typically applied to ignored speech. This research demonstrated the capability to detect aided cortical responses across different hearing loss degrees in the elderly for both attended and ignored stimuli. However, the significant neural tracking of the speech envelope was only observed at the group level. Moreover, the study was limited to the effects of linear HA amplification and slow compression, aiming to preserve the speech envelope.

Decruy et al. (2020) and Gillis et al. (2022a) explored the differences in neural tracking of speech envelopes between hearing-impaired and normal-hearing individuals, matching participants by age. Instead of typical HA use, these studies linearly amplified speech stimuli, ensuring 100% speech comprehension for hearing-impaired participants. Decruy et al. (2020) utilised backward modelling to reconstruct envelope accuracy, whereas Gillis et al. (2022a) employed a forward modelling approach. Both, studies reported enhanced neural tracking in hearing-

impaired individuals, although this difference was not statistically significant. One interpretation could be that adults with a hearing impairment exert more effort to distinguish target speech from background noise (Decruy et al., 2020, Verschueren et al., 2021). Another possible explanation is that a high-intensity stimulus was presented to hearing-impaired participants. However, Verschueren et al. (2021) tested this assumption by testing neural tracking across three intensity levels (30 dB, 60 dB, and 70 dB A), finding no significant effect. This suggests that amplification alone does not account for the enhancement in neural tracking. These findings highlight that hearing loss does not negatively impact neural tracking of continuous speech, supporting neural tracking as a valid method for assessing aided responses in hearing-impaired individuals. Nevertheless, the use of linear amplification in these studies does not accurately represent real-world HA use, which typically involves compression. This highlights the importance of investigating HA responses under more realistic amplification conditions. Specifically, future studies should focus on comparing pure linear amplification (uniform gain across frequencies) with compression-based amplification used in modern HAs, aiming to understand how these differing strategies affect neural tracking and auditory processing.

The distinction between aided and unaided neural tracking of speech envelopes has been explored in adults (Vanheusden et al., 2020) and children (Van Hirtum et al., 2023). Vanheusden et al. (2020) focused on subjects with mild to moderate hearing loss, most exhibiting a sloping high-frequency hearing loss profile. The stimulus in this study was low-pass filtered at 3000 Hz and presented at a level of 70 dB SPL. These parameters were chosen because of previous findings that frequencies above 3 kHz have no significant effect when measuring low-frequency cortical entrainment to speech (Vanheusden et al., 2020). The gain for the frequency range of the stimulus was set to 5 dB at low frequencies up to 1 kHz, then gradually increased to 30 dB at higher frequencies. The EEG data were collected via 32 channels and analysed using a linear decoder to reconstruct the stimulus envelope. The specific personal settings for the participant's hearing aids were not detailed. The study concluded that HA processing did not significantly change the correlation coefficient in backward modelling – a finding that reinforces the potential of using neural tracking of continuous speech in assessing aided responses. An important observation here is that the speech envelope was unlikely to have been affected by HA processing. Following the findings of Vanheusden et al. (2020), the absence of any significant effect of HAs may be attributed to normal hearing thresholds below 1 kHz, where the envelope is primarily shaped by low frequencies approximately between 2 Hz and 50 Hz (Rosen et al., 1992). It is possible to see a greater impact of HAs on the speech envelope with more gain and compression at lower frequencies. Since the stimulus envelope was the feature used for

measuring response presence, evaluating the impact of aiding on the envelope could provide deeper insights into how cortical responses are influenced by HA processing.

The morphology of cortical responses undergoes significant changes with age (Hall, 2015), highlighting the importance of investigating neural responses to continuous stimuli across different age groups. In a study by Van Hirtum et al. (2023), children with permanent hearing loss were assessed using neural envelope tracking to evaluate hearing aid effectiveness. The study included ten children aged 4 to 9 years, who listened to story segments presented at different levels under two conditions: aided and unaided. Each child used their personal hearing aids during the experiment. However, only six of the ten children completed the entire experiment, as younger participants often had shorter attention spans (Van Hirtum et al., 2023). This observation underscores the necessity of designing testing protocols with reasonable durations, particularly when assessing young participants. The study revealed a significant enhancement in neural tracking results with the use of hearing aids. Interestingly, this benefit was more pronounced at lower stimulus levels, where speech intelligibility is more challenging (Van Hirtum et al., 2023). It would be insightful to determine whether this finding is related to how the hearing aids process and potentially distort the stimulus envelope at different intensity levels. The isolated effect of the aided condition was not clearly measured in this study, as the comparison between aided and unaided conditions was conducted at the same presentation level, without accounting for the gain-induced increase in output intensity in the aided condition.

It is well established that the cortical auditory evoked potentials to repeated short stimuli are influenced by stimulus intensity. With both non-speech (Picton et al., 1970, Billings et al., 2007), and speech stimuli (Van Dun et al., 2012, Prakash et al., 2016), the peak amplitude increases, and the latency decreases as the stimulus intensity is raised. However, this effect shows saturation between moderate and high-intensity levels, particularly regarding amplitude (Prakash et al., 2016). Regarding neural tracking of continuous stimuli, only a few studies have explored the intensity effect (Ding and Simon, 2012a, Verschueren et al., 2021). Ding and Simon (2012a) examined the neural tracking of a speech stream in the presence of a competing speaker, varying the intensity level difference between the two streams by up to 8 dB. They discovered that the encoding of the speech envelope is independent of the intensity of individual speech streams, suggesting that the intensity difference investigated might have been too low to exhibit an influence. Verschueren et al., (2021) investigated various levels of stimulation, ranging from 20 dB to 80 dB A, where speech was presented in a quiet setting. The study employed both decoding (backward modelling) and encoding (forward modelling) methods in its analysis and included 20 adults with normal hearing. The EEG data were collected and analysed using a multichannel approach. It was observed that increasing the

intensity level did not affect the correlation value of the reconstruction. However, the effect of increasing the stimulation level was apparent in the TRFs, as the amplitude of the first peak significantly increased and the latency of the second peak decreased with higher intensity levels (see Figure 6.1). However, this effect was not observed between 60 dB and 70 dB A, aligning with the findings of Prakash et al. (2016) about the saturation effect at late responses. While Verschueren et al. (2021) identified specific effects of intensity changes on cortical responses to continuous stimuli, the sensitivity of a single-channel approach to these changes remains unexplored.

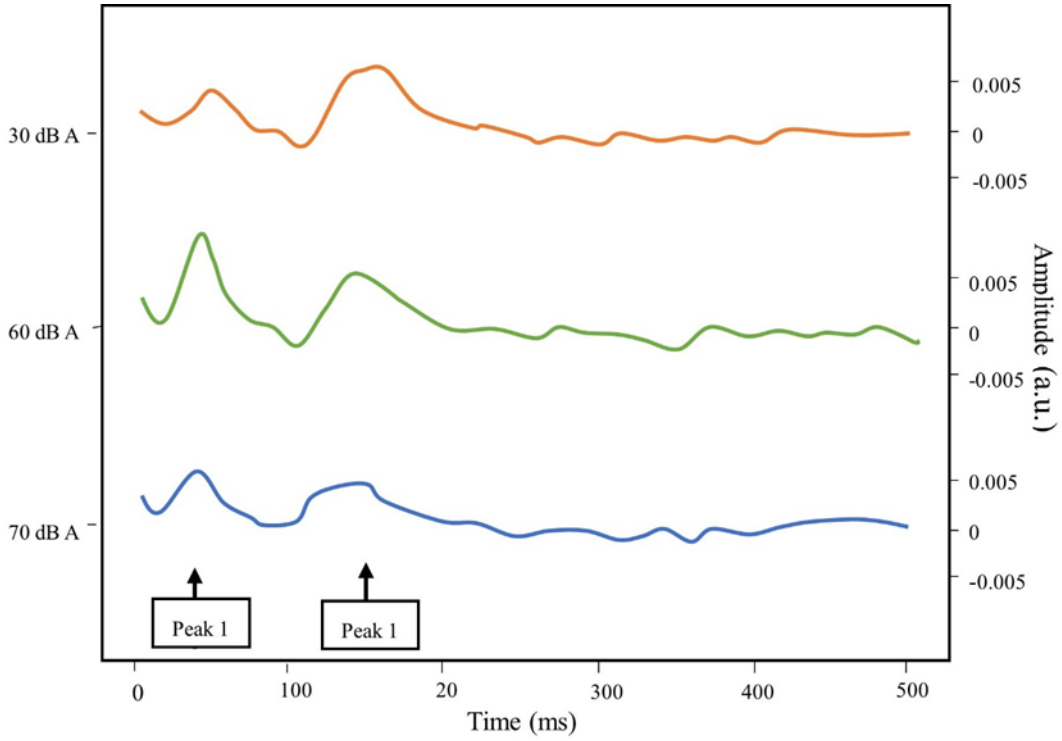

Figure 6.1 Result of the effect of intensity on peaks of averaged TRFs (Verschueren et al., 2021). *The figure illustrates the average centro-frontal TRF across participants for each intensity level, adapted from Verschueren et al., (2021)*

Another essential consideration in the analysis of neural tracking of aided responses is whether to use the audio envelope at the input of a HA microphone, or the output from the HA that goes to the ear, which may differ from the input signal; this factor was examined by Decruy et al. (2020). Their study found no significant difference in reconstruction accuracy between amplified and unamplified envelopes, noting that only linear amplification was applied, which would be expected to have a limited impact on the shape of the envelope. Mirkovic et al. (2019) explored the effects of HA processing on neural responses to ongoing speech, using outputs from a HA simulator. They argued that relying on the HA's input could lead to misleading results due to speech processing alterations. Nonetheless, employing unamplified envelopes would be a clinically more feasible approach, as recording HA outputs requires additional equipment. However, distortion in the acoustic envelope due to the HA processing could be a reason for

reduced neural tracking of the speech envelope. Therefore, further analysis is required to quantify this effect and determine if the envelope distortion is a significant confounding effect when measuring patients' neural tracking.

The type of speech or speech-like stimuli used for aided testing is a critical consideration, as HAs might respond differently to various inputs. Previous literature on tracking cortical responses to aided listening has employed intelligible speech stimuli. Nonetheless, in terms of the clinical applicability of speech tests, the use of universal or language-independent stimuli can be a superior option as it can be used across countries. The third study of this thesis assessed the effects of HA processing on both stimulus types (detailed in Chapter 5), revealing that speech-modulated noise received significantly less envelope distortion than actual speech (see Figure 5.7). Conversely, the ISTS exhibited envelope distortion levels similar to those of natural speech, suggesting that ISTS could more accurately represent how speech is processed through HAs. Additionally, ISTS may be subjectively closer to intelligible speech, as it includes both spectral and amplitude modulation, potentially making it more appealing than modulated noise. Moreover, ISTS can help reduce confounding factors related to comprehension differences among participants, such as language proficiency, cognitive abilities, and attention levels (e.g., native vs. non-native speakers, pre/post-lingual children).

The study aims to examine the impact of envelope distortion, induced by HAs, on cortical responses by comparing aided conditions with varying levels of distortion to the unaided condition. The purpose is to determine if envelope distortion affects the detectability of cortical responses. Additionally, the study explores the effect of increasing the stimulus intensity to provide insight into whether the HA processing effect on the stimulus envelope is caused by pure amplification or other signal processing. Finally, a non-language ISTS stimulus was used to determine if this produced robust cortical responses. This marks the first instance of measuring aided responses to ISTS using a single-channel recording. Although the future aim of this project is to test infants, the current stage investigates the feasibility of the test in normal-hearing adults. Results in adults will help in designing the future direction of the research.

6.1.1 Research questions

1. Does detectability of the cortical response to an ISTS stimulus change between aided and unaided conditions? Is this change related to the amount of envelope distortion?

2. Do the cortical responses to continuous ISTS significantly increase with stimulus intensity? This will provide insight into whether the hearing aid effect may be caused by pure amplification or other signal processing.

3. Is attention a significant factor in measuring cortical responses to aided responses in terms of detection sensitivity? If attention has no significant effect, it will help implement this test with infants.

4. When measuring responses to aided stimuli, is there a significant difference in detection sensitivity when using the original versus the processed envelope for correlation analysis? While using the original stimulus is more clinically feasible, better detectability with the processed envelope suggests a future research direction to record HA output for subjects in clinical settings.

6.2 Materials and Methods

The experimental approach of the current study, building on the findings detailed in Chapter 5, centred around EEG analyses using KEMAR-recorded stimuli. Focusing on the influence of HA settings, HA NR was deactivated. Deactivated NR was based on the finding of the previous study, which showed that NR has a different effect on envelope distortion across stimuli (see Figure 5.5). Furthermore, the research explored the role of compression in altering envelope distortion, employing various compression ratios to investigate their effects on cortical response measurements. This methodology provided an understanding of how different levels of envelope distortion impact speech-evoked responses.

6.2.1 Participants

This study initially involved 27 participants; however, due to EEG recording issues, data from two individuals were unusable. Additionally, one participant was excluded for hearing-related reasons, leaving 24 participants (16 females and 8 males) aged 18-40 years, with an average age of 29, for the final analysis. These individuals were recruited personally or via poster advertisements and received £20 each for their participation. Participants with incomplete or damaged EEG data were excluded to maintain data integrity. Ethics approval was granted by the University of Southampton Ethics Committee (ERGO: 52472), with all participants providing informed consent. Hearing sensitivity was confirmed as within normal limits through screening pure tone audiometry for each participant.

6.2.2 EEG recording

EEG data were recorded using a setup similar to that described in Chapter 4 (see Section 4.2.2 and Figure 4.1). For passive listening conditions, participants were asked to ignore the stimulus, while during active listening conditions, their focus was on the stimulus itself (detailed in

Section 6.2.3). In the passive listening scenarios, a series of silent documentaries with English subtitles were shown on a TV screen positioned in front of the participants, facilitating the passive engagement required for these conditions.

6.2.3 Stimulus and conditions

The primary stimulus in the current study was ISTS, which was presented in five different conditions. The presentation time for each condition lasted 15 min, aligning with the high detection rate of approximately 86%, as noted in the second study in Chapter 4 (see Figure 4.4). This duration was also deemed compatible with testing in infants. The stimulus was monaurally delivered to the right ear using a Focusrite sound card (Scarlett 18i8), and ER-3A inserted earphones. The monaural presentation mode was selected following ALR clinical protocols (Hall, 2015) and the second study design (Chapter 4). Two distinct ISTS stimuli, recorded via a microphone in KEMAR's ear canal, were employed: aided (with HAs) and unaided (without HAs). All stimuli levels were calibrated using a Bruel and Kjær Ear Simulator Type 4157.

The study included various conditions to assess auditory processing: aided at 50 dB LAeq and 70 dB LAeq, unaided at 50 dB LAeq and 70 dB LAeq, in which participants were instructed to avoid listening to the stimulus and focus on a silent documentary movie with subtitles presented on a screen. In addition to the four conditions, a fifth condition (unaided, 70 dB LAeq) involved active listening, where participants were asked to pay attention to the stimulus. In this condition, participants were instructed to mentally count clicks embedded in the stimulation sequence. These conditions were designed to investigate the effects of sound level, HA processing, and attention on auditory responses.

A comprehensive explanation of each condition and how stimuli are listed below. It is important to note that although a stimulus may be presented at a given level, such as 50 dB LAeq, the effects of amplification—whether through a hearing aid or via the ear canal of KEMAR—can result in a recorded level that may differ from the original (i.e., it may not be 50 dB LAeq). Additionally, the recorded stimulus could be played at various levels through the inserts presented to participants. Therefore, the recorded stimuli from KEMAR under different conditions (aided or unaided, and with various input stimulus levels) were adjusted to achieve the target stimulus levels presented to the subject through ER 3A insert phones.

1. Aided-50: ISTS was presented at 50 dB LAeq to KEMAR with HA, and the output was recorded and then adjusted to be at 70 dB LAeq level as an output from the inset.

2. Aided-70: ISTS was presented at 70 dB LAeq to KEMAR with an HA, and the output was recorded and then adjusted to match the 70 dB LAeq level. This was done to ensure the

output level was consistent with that of Aided 50 when presented to the listeners. The 70 dB LAeq input level resulted in significantly higher envelope distortion compared to 50 dB LAeq (see Section 5.3.2). Therefore, aided-50 and aided-70 were used to compare different levels of distortion without changing the sound level presented to the listeners. Also, as participants in the current study have normal hearing thresholds, the presentation level must be comfortable.

3. Unaided-50: ISTS was presented at 50 dB LAeq to KEMAR without an HA, and the recorded stimulus was then adjusted and calibrated to match a 50 dB LAeq output from the insets.

4. Unaided-70: ISTS was presented at 70 dB LAeq to KEMAR without an HA, and the recorded stimulus was then adjusted and calibrated to match a 70 dB LAeq output from the insets.

5. Unaided-70 (Attention): This condition involved repeating the unaided-70 scenario with active listening, where participants performed a counting task.

6.2.3.1 Repeated /da/ stimulus

The study included an additional condition focused on measuring CAEP in response to the repeated /da/ sound. This condition is crucial, as the design lacks a behavioral speech test, serving instead as a cross-check for the presence of a cortical response. Notably, cortical responses to repeated short stimuli have demonstrated a detectability rate of 95–100% in normal-hearing adults when presented above threshold levels (Munro et al., 2011). The absence of a behavioural speech test was based on the study's aim to use language-independent stimuli, as the subjects had different mother tongues. Averaged CAEPs are expected to yield robust cortical responses and a higher SNR. Participants who did not exhibit detectable CAEPs to /da/ were excluded. It is known from the second study (see Figure 4.4) that not all participants show detectable responses to continuous speech. Hence, measuring responses to a repeated /da/ sound was to ensure that absent responses were due to the subject simply having poor cortical responses, rather than a technical problem. Some individuals might have weak cortical responses even to repeated transit stimuli, as cortical responses are highly variable between subjects (Rothman, 1970). The responses to the /da/ sound were also used to explore correlations between the CAEP waveform and TRF peaks.

For the /da/ condition, the monosyllable /da/ stimulus was repeated 200 times with a 40 ms duration, presented monoaurally. The inter-stimulus interval was set at 1.11 s, and the stimulus was calibrated to 70 dB LAFmax (using a Type 4157, Bruel and Kjaer).

6.2.4 Data analysis

All analyses were performed offline using MATLAB 2022. The ISTS envelope was determined using the absolute value of the Hilbert transform, then downsampled from 44.1 kHz to 128 Hz with an appropriate anti-aliasing filter, followed by a 1 Hz - 30 Hz zero-phase filter. For EEG analysis, data from 32 channels were recorded, but the analysis was applied on either the vertex (Cz) or high forehead (Fz) channels. The EEG data was pre-processed: downsampled to 128 Hz, referenced to the mastoid electrodes (T7 and T8), and filtered between 1 Hz and 30 Hz to match the ISTS stimulus parameters.

Analysis methods included XCOR and TRF, using only the forward model. Both methods were explained in detail earlier in this thesis (see Section 3.3.1). Key analytical parameters were detection rate, detection time (see Section 3.3.2) and correlation values from both TRF and XCOR methods. The subject's response was considered present when a statistically significant correlation was detected according to the bootstrap method. Regarding detection, the bootstrap method was applied minute-by-minute to accumulating data, and a response was considered present if it was detected in three successive minutes or after 15 minutes of recording Detection time is defined as the recording time in minutes needed for each subject to show a significant correlation.

The presence of CAEP due to the /da/ sound was objectively evaluated using Hotelling's T2 statistics, as detailed by Golding et al. (2009). This approach entails automated detection of responses within EEG data, segmenting each data epoch into ten 50 ms bins across a 500 ms period following the stimulus. A Time-Voltage Mean (TVM) is computed for each bin, resulting in ten TVMs per epoch. The HT2 analysis then assesses whether any combination of these TVMs significantly deviates from zero. A *p*-value of 0.05 or lower indicates that a response is present. This method provides a rigorous statistical framework for identifying CAEPs in individual participants within the study.

6.2.4.1 Additional analysis

In the main analysis of aided conditions, the original ISTS envelope was used for the reasons mentioned in Section 6.1. An alternative analysis employed the processed envelope derived from the HA output. The study compared detection rate, and detection time between the original and processed envelopes. A significant difference in detection sensitivity when using the HA-processed envelope might indicate the need to explore methods for recording and utilising HA outputs if this technique is to be used in clinical settings.

6.2.5 Pilot study

Before the main experimental phase, a preliminary pilot study was conducted with four participants to evaluate the readiness of the equipment setup and to analyse the responses to the ISTS stimulus, which had not been previously used in neural tracking of continuous stimuli. EEG data from the Cz channel was collected, and the XCOR method was employed for analysis. The findings indicated that detectable responses were present in all four subjects after 15 min of recording under three different conditions: unaided-70 with both passive and active listening and aided-50. Notably, the aided-70 and unaided-50 conditions elicited significant responses in only three participants. The pilot data suggested that the ISTS stimulus provoked detectable responses in most tested conditions, and a duration of 15 min was sufficient to observe significant responses across all four subjects. An important observation was that the initial 20 min recording duration per condition proved to be excessively long, as it led to drowsiness in some participants by the end of the session. Importantly, existing literature indicates that cortical responses are influenced by the state of arousal (Hall, 2015). Therefore, for the main study, the duration of each condition was adjusted to 15 min to avoid fatigue and maintain participant alertness.

6.3 Results

The initial phase of the analysis was dedicated to examining the effects of three primary factors on cortical response detection to continuous speech-like stimuli: (1) HA processing (specifically envelope distortion); (2) sound intensity level; and (3) attention. This examination focused on two dependent variables: detection rate, and detection time. In the final analysis, 24 subjects were included for whom CAEP responses to /da/ measured from the Cz channel showed significant responses.

6.3.1 Detection rate

Figure 6.2 shows the detection rates for the five test conditions, utilising data from both Cz (blue) and Fz (orange) channels. Notably, the Cz channel exhibits a consistently higher detection rate across all five conditions and testing parameters, with a mean of 82.5%, compared to the Fz channel, which has a mean of 74.2 %. Given these findings, further analysis was concentrated on EEG data from the Cz channel for a more detailed comparison between conditions. As no significant difference in detection was found between the two methods of analysis, and the detection rate averaged across conditions was 83.3% for both methods, the XCOR analysis approach will be used in further analyses due to its simplicity and directness

compared to the TRF. Therefore, the XCOR results will be included in the detection rate, detection time, and correlation value analyses.

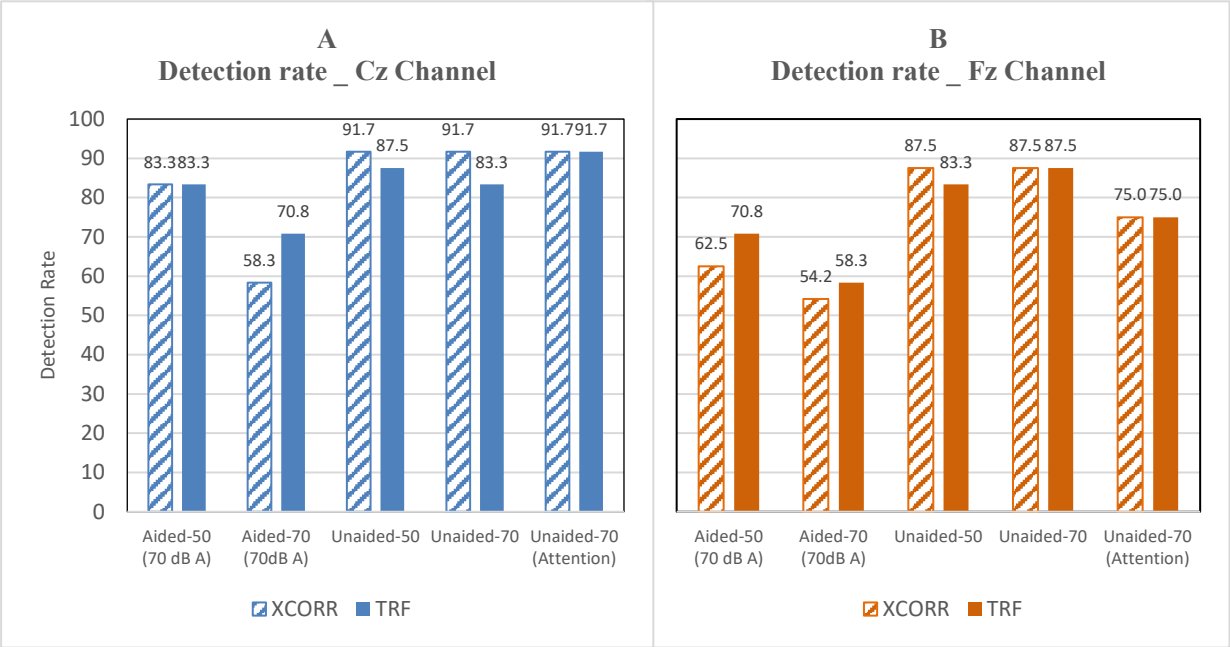

Figure 6.2 The detection rates of cortical responses measured from the Cz and Fz channels using XCOR and TRF analyses. *(A) Cz channel and (B) Fz channel. The first two conditions from the left were aided with two input levels to the HA at 50 dB LAeq and 70 dB LAeq, but the output levels have been scaled to all be 70 dB LAeq. The third and fourth columns are unaided conditions with presentation levels of 50 dB LAeq and 70 dB LAeq. The rightmost column depicts the attended condition, where subjects were instructed to count clicks in the stimulus for unaided 70 dB LAeq stimulation.*

Figure 6.2 (A) shows that the higher detectability for the aided condition occurs at low envelope distortion aided-50, at 83.3% (20 out of 24). Detectability decreases with aided-70, which has a higher envelope distortion (see Section 5.3.3), falling to 58.3% or (14 out of 24) for XCOR.

McNemar tests were conducted to assess the significance of changes in detection rate across conditions. To assess the impact of HA envelope processing on detectability, unaided-70 was compared to aided-50 and aided-70 conditions with intensity levels adjusted to 70 dB LAeq for all three. A Bonferroni correction was applied to account for multiple comparisons, adjusting the alpha level to 0.025 (0.05/2). A significant difference was noted when comparing the detection rates of unaided-70 (91.7%; 22 out of 24) to aided-70 (58.3%; 14 out of 24), with $p < 0.01$, indicating a higher detection rate in the unaided-70 condition. This suggests that distortion of the speech envelope by HAs at 70 dB LAeq input level (when the aid compresses the most) reduces response detection.

Furthermore, the investigation extended to the impact of intensity level on detection by comparing rates in unaided-70 and unaided-50 conditions. This comparison, conducted via the

McNemar test, revealed no significant difference, indicating that the intensity level alone did not affect detectability ($p > 0.05$).

In addition, the analysis of detection rates in attended versus unattended conditions within the unaided-70 settings did not show any significant differences, as determined by the McNemar test. This suggests that attentional factors, in the context of this study's settings and conditions, did not significantly influence detection rates.

6.3.2 Detection time

Figure 6.3 displays the detection rate per minute of recording time for all tested conditions, as measured from the Cz channel across all subjects. The expected false-positive rates from the binomial distribution under a no-stimuli condition are between 0 and 4 (at a significance level of 5%). Across all testing conditions, false-positive rates for both stimuli remained within the expected range.

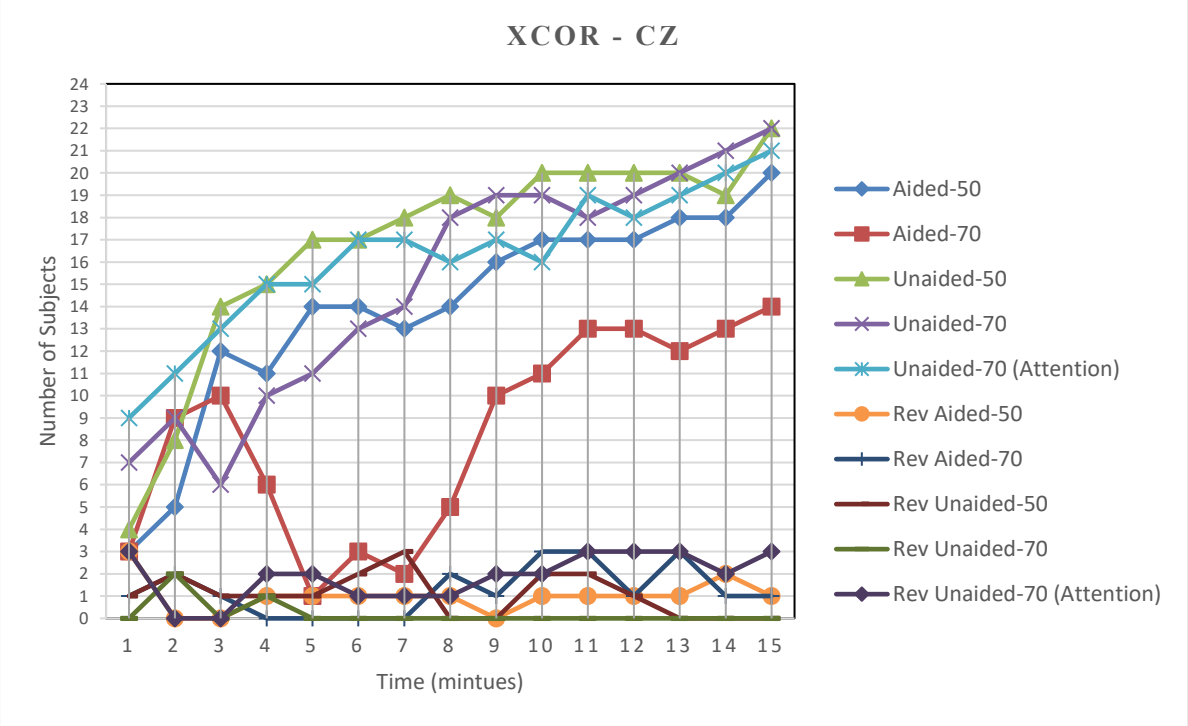

Figure 6.3 The number of significant cortical responses by minute for each test condition measured from Cz. *Detection thresholds at p = 0.05 were found from bootstrapping of non-synchronous data. The x-axis represents the recording time ranging from 1 min to 15 min. The y-axis presents the number of subjects with significant responses. A reversed condition (indicated as Rev) was measured for each tested condition where the stimulus was inverted to assess the false positive rate.*

Figure 6.4 presents the detection times across the five experimental conditions. The Shapiro-Wilk test indicated a significant deviation from normal distribution for all conditions with p-values less than 0.05, necessitating the use of non-parametric tests for further analysis.

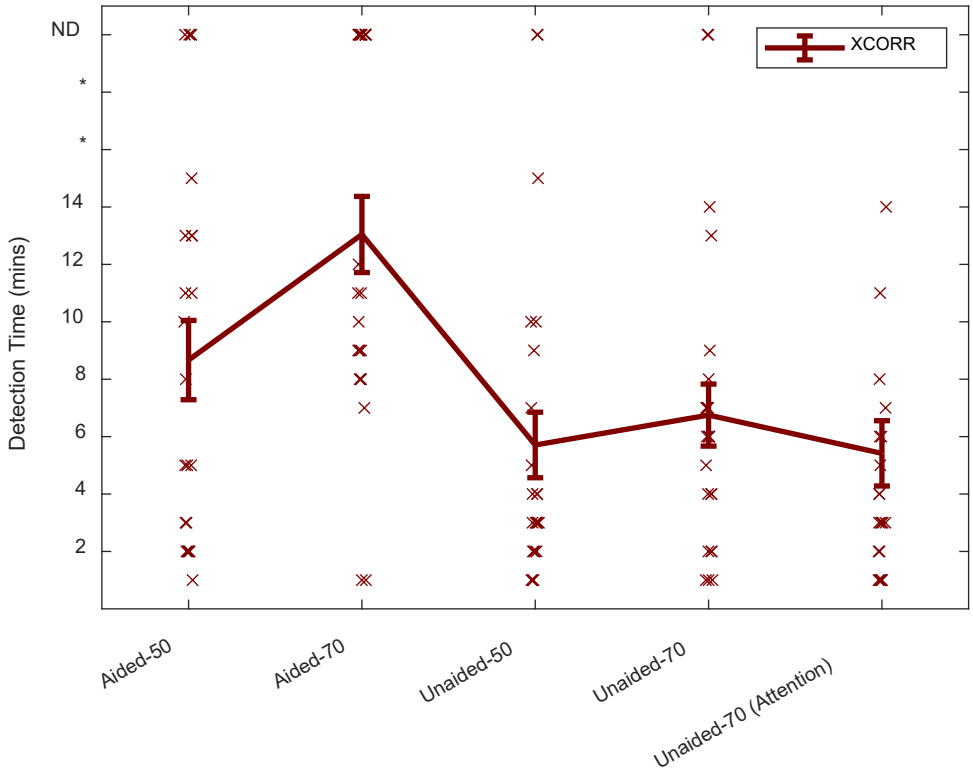

Figure 6.4 Mean and individual detection times. *The solid red lines indicate mean XCOR detection time, with error bars indicating the mean +/- 1 SE. Individual subject measurements are shown as red crosses. The y-axis represents the detection time in minutes, and ND indicates no detection (within the duration of the recording). The initial two columns from the left represent the aided condition, each with varying degrees of envelope distortion: the aided-50 condition exhibits low envelope distortion, while the aided-70 condition shows high envelope distortion. The third and fourth columns illustrate unaided conditions with different presentation levels (50 dB LAaq and 70 dB LAeq, respectively). The column on the far right demonstrates the attended condition, in which subjects were asked to count clicks in the stimulus during the unaided 70 dB LAeq stimulation.*

It should be noted that for cases with no detection, a detection time of 20 min was assigned. The analysis of the XCOR method revealed statistically significant differences between experimental conditions ($p < 0.001$) according to the Friedman test. Post-hoc Wilcoxon Signed Ranks tests were conducted to explore the effects of envelope distortion, intensity levels, and attention on detection time. A Bonferroni correction was applied to account for multiple comparisons, resulting in a corrected significance threshold of $p < 0.01$. Results showed a significant difference between aided-70 (high envelope distortion) and unaided-70. The aided-70 condition (13.04 ± 6.5 SD minutes) had a significantly longer detection time compared to the unaided-70 condition (6.7 ± 5.3 SD minutes) ($p = 0.001$). No significant effect of intensity level or attention was detected.

Figure 6.5 details individual XCOR detection times, highlighting two notable observations. There is considerable variation in detection time data. However, some subjects appear to have

consistently low detection times (e.g., S14) and some consistently high (e.g., S24). Secondly, the consistent detection time tends to be consistent across conditions for each subject. For example, S14 and S27 exhibit fast detection (1 min - 3 min) across all conditions, while S13 and S24 are slow (5 min – 20 min). However, there is also some wide variability within individuals (e.g., S06 and S21). This observed variability led to further analysis to determine if the detection time was significantly different between subjects.

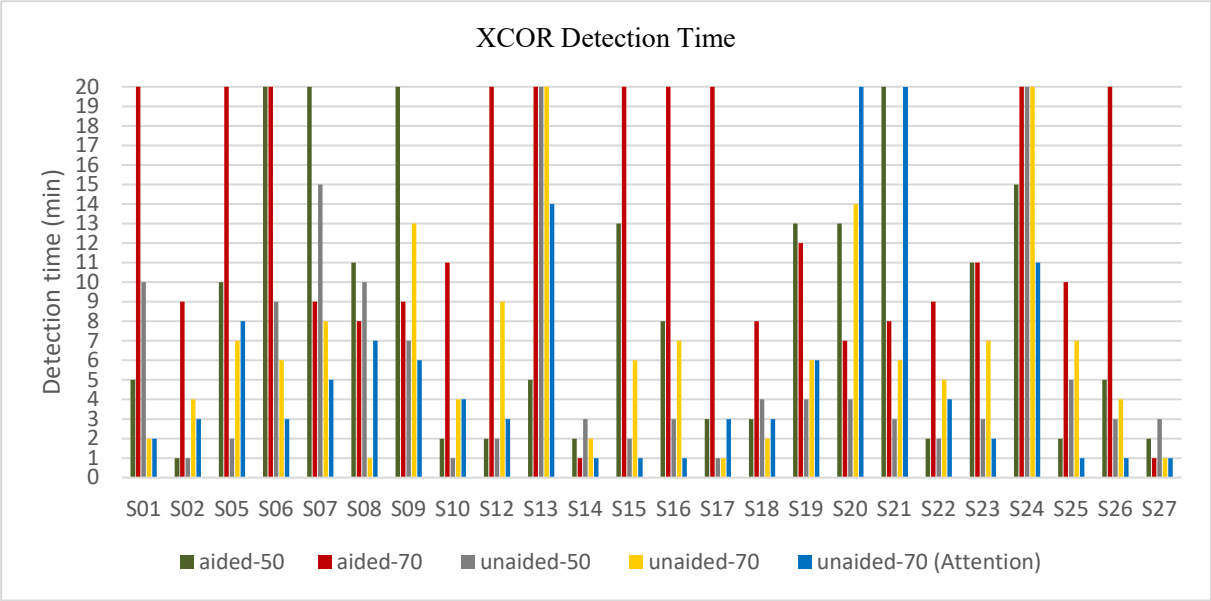

Figure 6.5 Individual XCOR detection Time. *This figure represents the XCOR individual detection time for each subject in the tested conditions. Each condition is represented in colour: aided-50 (green), aided-70 (red), unaided-50 (grey), unaided-70 (yellow), and unaided (attention) (blue). A value of 20 was assigned to conditions where no detection occurred.*

Figure 6.6 displays error bars for the mean and SE of the five conditions for each subject. To test if there were significant differences between subjects, the detection time data was transposed so that the original number of subjects (N = 24) became the number of conditions, and the original number of conditions (n = 5) became the number of subjects. The results of the Friedman test showed a significant between-subject effect (p = 0.001). Post-hoc pairwise comparisons using the Wilcoxon Signed Ranks Test were conducted to assess differences across the 24 subjects. To adjust for multiple comparisons, a Bonferroni correction was applied, setting the significance threshold at 0.00018 (original alpha = 0.05). Under this stringent criterion, no significant differences in pairwise comparisons between subjects were detected.

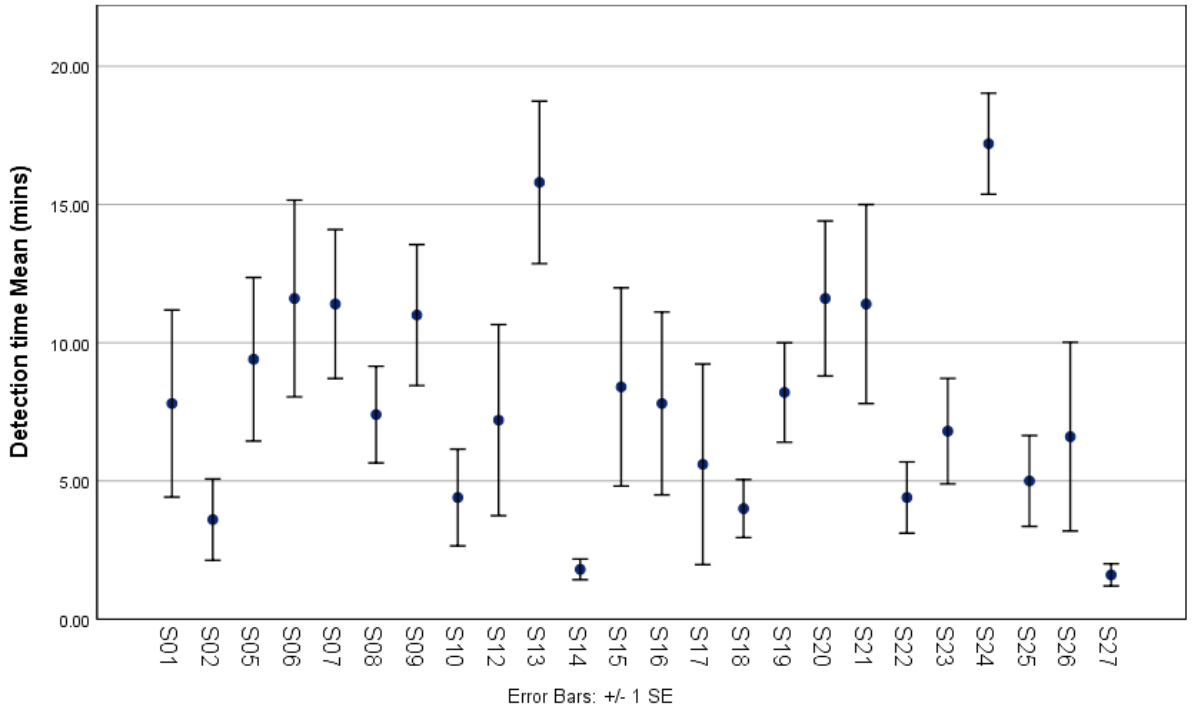

Figure 6.6 Detection time error bars of each subject. *The figure illustrates the mean and standard error (SE) of the detection time across the five conditions for each subject, as shown in Figure 6.6.*

6.3.3 Effect of stimulus intensity on TRF

Previous results in this chapter of detection rate (see Section 6.3.1, and detection time (see Section 6.3.2), revealed no significant differences with increasing intensity. An additional analysis was conducted to compare the peak amplitudes from the TRF functions between the unaided-50 and unaided-70 conditions. Figure 6.7 shows the group average of the TRFs across participants for these two conditions unaided-50 and unaided-70, along with individual TRF variability. It is clear from the figure that a positive peak occurs at latency between 70 ms and 125 ms, and a negative peak occurs between 125 ms and 200 ms. The peak amplitude of the two peaks was compared between the two conditions.

The Shapiro-Wilk tests indicated that individual TRF peak amplitude values did not significantly deviate from normality ($p > 0.05$). However, this does not confirm that the data is normally distributed, only that there is insufficient evidence to reject the assumption of normality. The results of the paired sample t-test showed no significant difference between either the positive or the negative peaks of the TRF across the two conditions. These findings suggest that, with the current sample, intensity did not significantly affect the peak amplitude of the TRF.

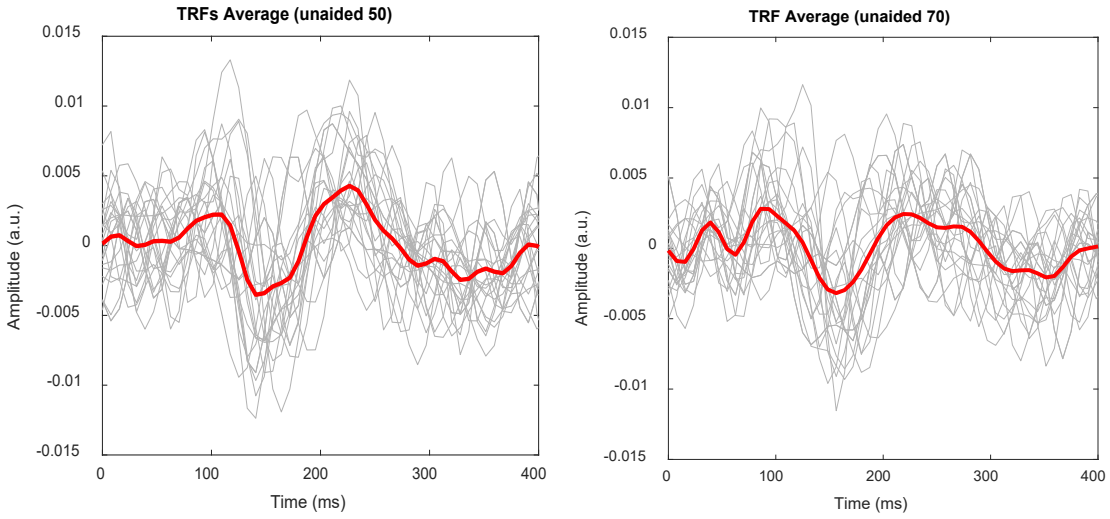

Figure 6.7 The grand average of TRFs at different intensity levels. *Red solid lines show averaged TRFs measured from the Cz channel, and individual TRFs are shown as grey lines. The left panel displays the unaided-50 condition, while the right panel shows the unaided-70 condition.*

6.3.4 The impact of using the speech envelope at the output of the hearing aid versus the original speech envelope on response detection

The final question addressed in this study pertains to the impact of employing the processed envelope — derived from the HA output — in the aided conditions, as opposed to the original stimulus. This section presents the detectability and detection time based on the analyses conducted earlier. Both aided conditions, aided-50 and aided-70, were examined in this context.

Figure 6.8 shows the detection rates after 15 min of recording for the aided conditions, comparing the use of the processed envelope (green bars) versus the original envelope (red hashed bars). A review of the data shows a trend towards higher detection rates with the processed envelope. Therefore, McNemar tests were conducted to evaluate the significance of this difference between the processed and original envelopes across the two aided conditions. To adjust for multiple comparisons, a Bonferroni correction was applied, setting the alpha level to 0.025 (0.05/2). The tests revealed no significant differences with aided 50 ($p = 0.063$) and with aided 70 ($p = 0.031$), indicating that the use of the processed envelope does not significantly affect detectability. However, without the Bonferroni correction, a significant increase in the detection rate was seen in the XCOR aided-70 conditions with the processed envelope (83.3%) compared to the original envelope (58.3%), $p < 0.05$.

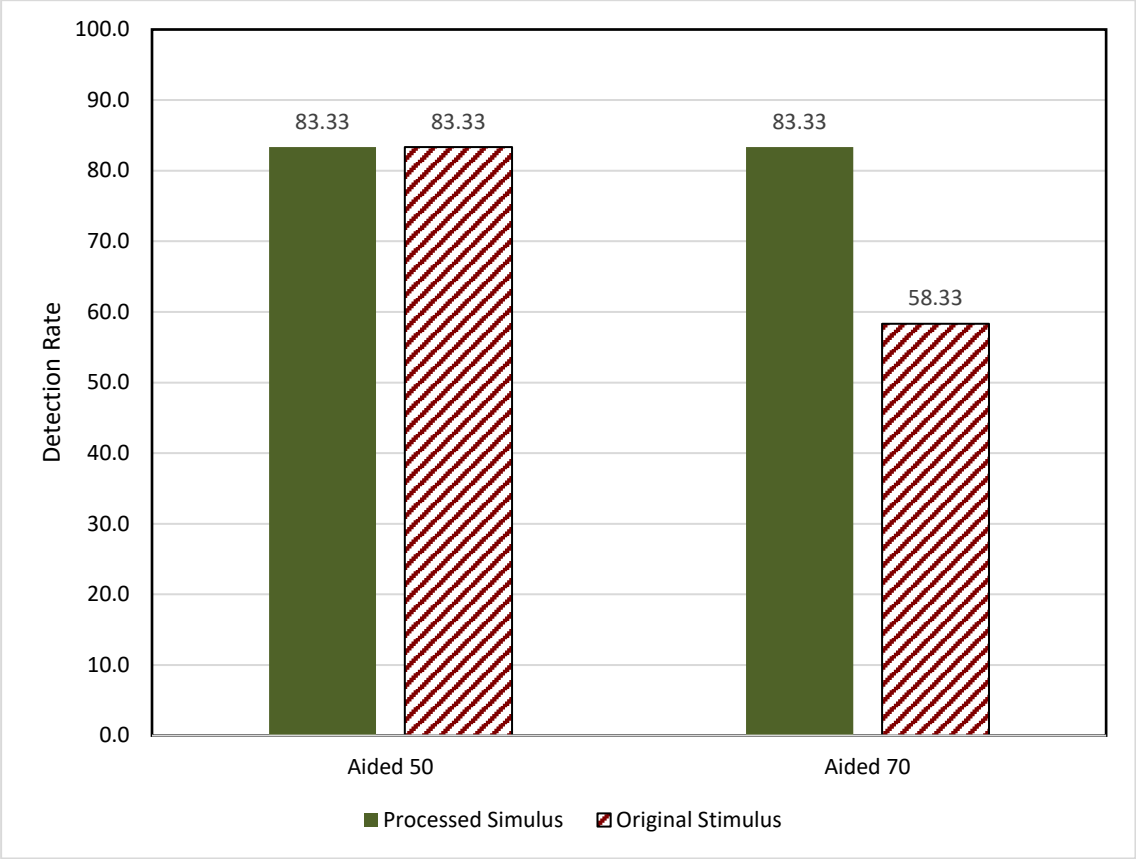

Figure 6.8 Detection rates using either the unprocessed ISTS envelope or the processed ISTS envelope in the aided conditions. *The detection rate represents the percentage of subjects with significant cortical responses (from 24 subjects). The columns on the left show the aided-50 condition with a low envelope distortion level. The columns on the right show the aided-70 conditions with a high envelope distortion level. The solid fill column represents the analysis result when using the stimulus recorded from the HA output, while the shaded column represents the original stimulus results.*

Figure 6.9 shows the mean detection time of the aided conditions when using the original and processed envelopes. It is clear from the figure that the mean detection time is longer when using the original envelope compared to the processed envelope in the aided-70, which is the condition with higher distortion. A repeated measures analysis using the Friedman tests was conducted to assess the significance of this effect. The test shows that the processed stimulus did not significantly affect the detection time $p > 0.05$. Since a strong effect was expected in the aided-70 condition, as indicated by the detectability results in the previous paragraph (see Figure 6.8), a Wilcoxon test of comparison was made. The results of the Wilcoxon test comparing the detection time of the original and processed envelopes in the aided-70 condition showed no significant difference ($p = 0.075$).

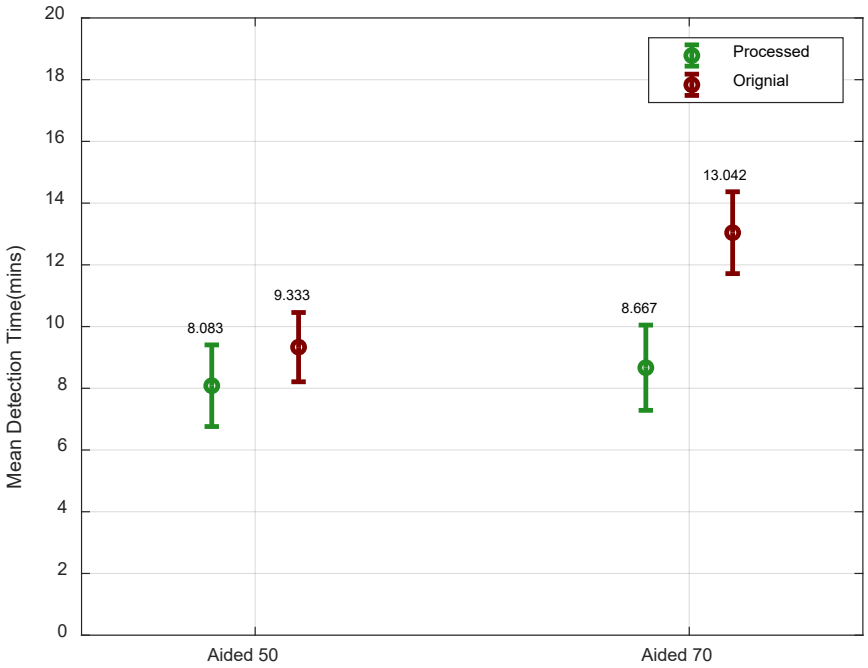

Figure 6.9 The mean detection time of the aided conditions, comparing the results of using the original envelope to the processed envelope*. The error bars indicate the SE of the mean. The y-axis represents the detection time in minutes. The green error bars show the results of using the processed envelope, while the maroon error bars show the results of using the original envelope. The left bars show the results of the aided-50 condition and the right bars show the aided-70 condition.*

6.4 Discussion

This study explored cortical responses to continuous ISTS stimuli under aided and unaided conditions. Specifically, the effects of envelope distortion, intensity, and attention on detection rates and detection times were explored. Notably, the speech processing carried out by HAs significantly influenced cortical responses. Aided conditions exhibited lower detection rates and required longer times for detecting responses compared to unaided conditions at the same presentation level. This impact was particularly pronounced in high envelope distortion situations (aided-70). Additionally, our findings indicate that neither increasing the stimulus intensity by 20 dB nor altering the focus of attention affected cortical responses. Interestingly, under high envelope distortion conditions, utilising the processed envelope for measuring cortical responses significantly improved detectability. This suggests that future research should focus on finding a clinically feasible approach to using HA output in clinical settings to analyse cortical responses to continuous speech.

6.4.1 Aided detectability and effect of envelope distortion

Evaluating the clinical feasibility of any tool involves assessing its sensitivity or detection rate. In this study, detectability was 83.3% in the aided condition with low envelope distortion, whilst it dropped to 58.3 % in conditions with high envelope distortion. Previous research on aided responses to continuous stimuli has typically focused on the group results, often overlooking individual detectability (Petersen et al., 2017, Vanheusden et al., 2020, Decruy et al., 2020, Gillis et al., 2022a). To bridge this gap, figures from existing literature that detailed individual results were reviewed. This approach allowed for a comparison between the current study's findings and previous research. For instance, Vanheusden et al. (2020) found that one out of 17 subjects showed no significant responses in the aided condition.

Furthermore, Van Hirtum et al. (2023) explored the use of neural tracking of speech envelope in assessing the HA benefit in children using backward modelling. The correlation values were compared between aided and unaided conditions under five presentation levels ranging from 30 dB A to 70 dB A. They reported a detectability of 100% in the aided condition across various stimulation levels for eight children included in their final experiment. The detectability of aided responses in the current study is lower than previous literature (Vanheusden et al., 2020, Van Hirtum et al., 2023). One possible explanation is the analysis approach, while both studies use a linear decoder which reconstructs speech envelopes from EEG compared to the cross-correlation approach implemented here. Additionally, the single-channel approach implemented here contrasts with the multichannel analysis used in prior studies.

The choice of the single-channel approach was mainly based on the project's overarching aim: to assess the clinical feasibility of aided cortical responses to continuous speech stimuli. For the implementation of a clinical tool, the sensitivity in detecting significant responses in normal-hearing individuals should be close to 100%, a benchmark not achieved in this study. High test sensitivity is crucial to reduce the risk of false outcomes in clinical use (Bright et al., 2011). However, the sensitivity results from the current study suggest that the single-channel analysis employed may not be sufficient for clinically assessing aided cortical responses to continuous speech stimuli. In addition, the stimulus used in the current study was ISTS, which might have contributed to the lower performance in detectability. Figure 6.10 shows a compression between natural speech and ISTS in silent pause duration within the amplitude envelope, with natural speech having longer pauses. Deoisres et al. (2023) found that speech with additional long pauses produces a significantly higher correlation than natural speech.

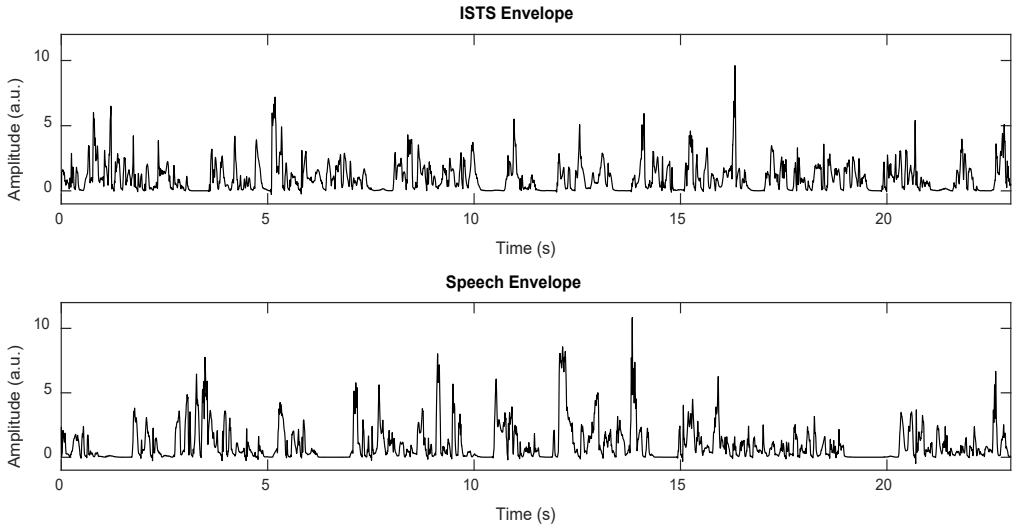

Figure 6.10 Comparison of ISTS and speech envelopes in gaps duration within the amplitude envelope. *This figure shows the difference in pause duration between the ISTS and natural speech stimuli. The speech stimulus exhibits longer silent pauses compared to the ISTS, highlighting a key distinction in their temporal structures.*

The aided condition with high envelope distortion demonstrated a significantly lower detection rate, and longer detection time compared to the unaided condition, diverging from previous studies that compared aided and unaided neural tracking of the speech envelope. For instance, Vanheusden et al. (2020) reported no difference, while Van Hirtum et al. (2023) observed an enhancement in the aided condition. Several factors might explain this discrepancy. Firstly, our study employed single-channel and forward modelling techniques, as mentioned earlier. Another difference lies in the participant groups; our study involved individuals with normal hearing who might find processed sounds unfamiliar, unlike participants with hearing loss in previous studies, who were accustomed to HAs.

Moreover, the envelope distortion of the stimuli in our study was measured before the experiment began (Chapter 5), which allowed us to investigate how different envelope distortion values affect the measured cortical responses. In contrast, the specific amount of envelope distortion was not detailed in the previous studies. It is possible that the distortion in our study, especially in the high-distortion condition (aided-70), was greater than that in previous studies. In Vanheusden et al. (2020), the hearing threshold at low frequencies was within normal limits, suggesting minimal distortion of the speech envelope. Additionally, both studies used personal HAs, which could have resulted in higher detection as the signal was familiar to the participants.

Finally, the method of stimulus presentation could also contribute to the observed differences. Previous studies used free field stimulation which is required with HA users, allowing for

binaural hearing, which might enhance loudness perception due to 'binaural loudness summation'—a phenomenon where a sound presented to both ears is perceived as louder than if presented to just one ear (Moore and Glasberg, 2007). In contrast, our study presented the stimulus monaurally via inserts, potentially impacting perceived loudness and detectability.

6.4.2 Effect of stimulus intensity

Examining how increases in intensity affect aided cortical responses to continuous speech stimuli could offer deeper insights into evaluating the benefits of HAs. It has been shown that neural tracking of the continuous speech envelope using backward modelling exhibits higher reconstruction quality or correlation as intelligibility increases (Ding and Simon, 2013, Vanthornhout et al., 2018). Therefore, measuring a higher correlation with increasing intensity might indicate improved speech detection. The current study investigated the effects of a 20 dB increase in intensity, finding no significant changes in response detection rate, and detection time.

Our results align with those of Verschueren et al. (2021), who reported no significant difference in correlation values of backward modelling between 30 dB A, 60 dB A, and 70 dB A stimulation levels. In contrast, Van Hirtum et al. (2023) reported a significant increase in reconstruction accuracy with increasing the level of stimulation as they tested at 40 dB, 50 dB, 60 dB and 80 dB. The difference was more noticeable between low level (40 dB) and high level (80 dB). In the context of the current study, the quietest level was moderately loud for normal hearing participants (50 dB LAeq), which might result in cortical response structuration. One limitation of the current study was the limited range of intensity tested. This was due to the time taken for the experiment to investigate other factors, such as HA processing effects and attention effects. Future work could investigate various levels, ranging from 0 dB SL , as in this study the low level (unaided-50) was approximately 25 dB SL for normal hearing individuals.

Regarding the amplitude peak of the TRF function, no significant difference was observed in the two main peaks investigated (70 ms - 120 ms and 125 ms - 200 ms) when increasing stimulus level by 20 dB. These findings are inconsistent with prior studies (Verschueren et al., 2021, Van Hirtum et al., 2023). Verschueren et al. (2021) identified significant peaks at latencies of 20 ms - 70 ms and 100 ms - 200 ms, observing a significant increase in the first peak when the stimulus increased from 30 dB A to 60 dB A. It is possible that the effect of intensity is only seen between low SL (near threshold) and moderate levels and not between moderately high levels as in the current study. Van Hirtum et al. (2023) reported similar findings, with significant increases in the first peak (50 ms - 100 ms) with increasing intensity. A potential explanation for the intensity effects observed in previous studies is using averaged TRFs across frontocentral channels.

Averaging several TRFs can accentuate the peak in the function. However, employing multichannel analysis may compromise clinical feasibility. Our findings show that with single-channel analysis, the change from moderate to loud stimulation showed no significant effect on measuring evoked responses to continuous speech.

6.4.3 Effect of attention

The results of this study indicate that attention did not significantly influence cortical responses to ISTS. Detectability (see Figure 6.2) and detection time (see Figure 6.4) remained non-significantly changed when the listening condition shifted from passive to active. Generally, auditory cortical responses are modulated by attention (Picton and Hillyard, 1974, Fritz et al., 2007, Näätänen et al., 2011). However, studies focusing on neural tracking of the speech envelope in conditions of quiet and normal hearing found no significant difference between active and passive listening (Kong et al., 2014, Vanthornhout et al., 2019), which is consistent with our results.

The nature of the stimulus used, and the attention task here might contribute to these results. The unintelligible nature of the ISTS may affect the ability to sustain attention. It is important to consider the required task and the focus of attention when examining its effects on neural responses (Fritz et al., 2007). In studies involving continuous intelligible speech, tasks typically require participants to focus on a speech stream against background noise or competing speaker, followed by questions or subjective estimations of understanding (Ding and Simon, 2012a, Kong et al., 2014, Vanthornhout et al., 2019, Gillis et al., 2022a). Research has shown that neural tracking of the attended stream is enhanced compared to the ignored stream (Ding and Simon, 2012a, Kong et al., 2014, Wang et al., 2020, Mirkovic et al., 2019). In this study, the task involved counting clicks embedded within the stimuli, suggesting that the clicks represented the attended stream, while the ISTS formed the ignored stream. This setup provides a plausible explanation for the lack of an attention effect observed in tracking the ISTS envelope. Notably, the absence of a dependency on attention may be advantageous in applications involving children, where reliably controlling attention can be challenging (Roebuck and Barry, 2018).

6.4.4 Comparing the original and the processed envelope in aided conditions

One of the objectives of this project is to investigate the clinical feasibility of detecting cortical responses to aided continuous speech. Comparing the original and processed envelopes of the HA was crucial for this purpose, as it might affect the measured response parameters. The results demonstrated no significant difference with including both aided conditions. However,

including the results where a significant effect is expected (aided-70), detectability significantly rose from 14 to 20 out of 24 using the XCOR analysis method (see Figure 6.8). It is also interesting to note that, for both envelopes, the phonetic content is the same, but the acoustics are different. This indicates that the response primarily reflects the acoustic stimulus rather than being specific to the language content.

Decruy et al. (2020), reported no significant difference in reconstruction accuracy between the original and amplified stimulus, similar to the current result with aided-50. However, given the improvement in detectability evident in conditions with high envelope distortion, using the processed envelope for clinical application should be considered. This might be an approach to improving the detectability of cortical responses to aided continuous speech in future research. It is also crucial to highlight that a reduction in correlation can be expected with HA processing, especially with higher compression, regardless of the impact on speech perception. Another approach to consider for clinical use is to avoid testing with the HA in a condition that will introduce significant distortion.

6.4.5 Cortical responses to continuous ISTS stimulus

To the best of our knowledge, this is the first study to employ the ISTS stimulus for neural tracking of the speech envelope, demonstrating significant potential for clinical feasibility due to its language independence and similarity to speech regarding HA processing (Chapter 5). The detectability of unaided conditions was explored to examine the detectability of ISTS without the effect of HA processing. The unaided-70 condition, with passive listening detectability results from the current study, can be considered promising for future clinical applications. The detection rate was over 91% after 15 minutes of recording, with a mean detection time of 6.7 min. The fact that the ISTS, despite being unintelligible, showed high detectability highlights that the assessment in the current study targets speech detection, which is the initial stage of speech perception. Speech perception also involves three additional stages: discrimination, identification, and comprehension (Erber, 1976). The ISTS could be a superior option for clinical testing, offering a universally applicable stimulus across various language backgrounds. Additionally, its lack of comprehensibility might eliminate confounding cognitive factors from the analysis, such as those affecting cognitively impaired individuals, children, or variations in attention.

6.5 Conclusion

This chapter explored the feasibility of using single-channel analysis to measure cortical responses to aided continuous ISTS stimuli. The detectability of aided responses was found to

be less than 100% (aided-50 = 83.3% and aided-70 = 58.3 %), suggesting overall limited clinical feasibility for single-channel measurements. Envelope distortion introduced by HA processing had a significant impact on the measured responses. Conditions with higher distortion levels were associated with lower detection rates (58.3%) compared to the unaided condition (91.7%), and longer mean detection times (13.04 min) compared to (6.7 min). Nevertheless, response detection with ISTS in the unaided condition showed detectability of > 91, which could be reasonable for clinical use. This indicates that ISTS could be promising for clinical utility due to its applicability across various language speakers.

Additionally, the study observed no effect of attention in cortical responses to continuous ISTS, an advantageous property when testing infants, where controlling attention is challenging. Using processed envelopes instead of original ones significantly improved the detection rate, especially under high envelope distortion conditions. One direction for future research is to find a clinically feasible method to use HA output to detect cortical responses to continuous stimuli. While this research demonstrated some aspects of the clinical feasibility of using cortical responses to aided continuous ISTS, improvements in detectability are necessary. Possibly, more powerful multichannel analysis may be needed for detection at this stage.

Chapter 7 General Discussion and Conclusion

The primary objective of this thesis was to investigate the clinical feasibility of using cortical responses to continuous speech stimuli as a means for objective speech detection via HAs. The overarching aim was to develop a clinical test for evaluating aided speech detection in infants. To achieve the thesis objective, a series of interconnected studies were carefully designed, exploring important factors for infant testing in clinical settings. This final chapter aims to provide a cohesive understanding of the topic by discussing and linking the key findings from each study.

The exploration began by assessing the feasibility of using a single channel for neural tracking of continuous speech (Chapter 3). This approach is crucial due to its clinical availability and reduced preparation time. Next, the thesis investigated detection rates of neural responses during passive listening and with unintelligible speech stimuli (Chapter 4), as controlling attention in infants is challenging, and non-language-specific stimuli provide a universal testing approach.

The focus then shifted to examining the use of HA-processed stimuli through two distinct studies. The first study explored the effects of different HA settings on multiple speech stimuli by measuring envelope distortion (Chapter 5), which results in varied HA outputs with differing levels of distortion. The second study involved EEG measurements in response to the pre-recorded unaided and aided stimuli with varying levels of distortion (Chapter 6), aimed at determining whether higher levels of envelope distortion impacted the detectability of cortical responses. Each study was designed to build upon the insights gained from its predecessors, progressively deepening the understanding of the subject. Below are the principal questions that guided these studies are outlined:

1. Using existing methods for measuring cortical responses to continuous speech (XCOR and TRF), what are the detection rates and detection times required when measuring responses using a single channel, either Cz or Fz? (Chapter 3)

2. What are the detection rates and detection times of cortical responses to continuous speech using data from a single channel during passive listening scenarios (without active attention)? (Chapter 4)

3. Is there a significant effect of speech stimulus intelligibility on detecting cortical responses to continuous speech? (Chapter 4)

4. Is there a significant difference in HA-induced envelope distortion between speech and speech-like stimuli (natural speech, speech-modulated noise, and ISTS)? Additionally, how does changing the HA compression ratio or input level affect envelope distortion? (Chapter 5)

5. Is there a significant difference in the detection rates and detection times of cortical responses to aided versus unaided continuous ISTS stimuli? (Chapter 6)

The first section of this chapter will discuss the two analysis approaches used in the current project. The following section will discuss the different aspects of clinical feasibility explored in this project. These aspects include detectability, detection time, stimulus type, and effects of attention. Each aspect will be linked to the studies conducted in this project, comparing the results and highlighting how the findings from each study influenced the design of subsequent studies. Additionally, these aspects will be linked to the project's research questions. The chapter will conclude with a discussion of limitations and suggestions for future work.

7.1 Approaches to Analysis

The first study in the project compared the detection sensitivity and mean detection time between two main analysis approaches, XCOR and TRF. The forward TRF method included not only the significance of the correlation values but also the peak and power of the TRF (see Section 3.2.1). To the author's knowledge, the detectability of cortical responses to continuous speech using the peak and the power of the TRF impulse response as parameters has not been investigated. The results of the first study clearly showed that the significance of the correlation, whether using XCOR or TRF, would be better suited for clinical use, as both demonstrated a 100% detection rate compared to 82% with TRF Power and 76% with TRF Peak. Additionally, significantly lower mean detection rates were observed with both XCOR and TRF-COR compared to TRF Peak and Power. No significant difference was found between XCOR and TRF-COR in both detection rate and time.

Both XCOR and TRF-COR were also used in the second study (Chapter 4). The results showed no significant differences in detection rate or mean detection time for either stimulus (speech and speech-modulated noise). In the fourth study (Chapter 6), both analysis approaches were implemented. However, as they didn’t show a significant difference in the results, only XCOR was used to compare the different conditions in that study. Based on the findings from the three studies, and to ensure consistent comparison, the outcomes of XCOR will be considered when

comparing the findings between the studies in this thesis (Section 7.2), as no significant differences were observed between the analysis methods.

7.2 Clinical Feasibility Factors

7.2.1 Single channel detectability

The first research question related to the feasibility of using a single measurement channel will be discussed. To the best of our knowledge, the detectability of auditory evoked responses to continuous speech stimuli has only been implemented using multichannel EEG recording systems, with a minimum of 32 channels in adults (Vanheusden et al., 2020) and 10 channels in infants (Ortiz Barajas et al., 2021). Current audiology clinical settings for AEP tests typically use single-channel recordings (a 3-electrode montage) with the positive electrode on the high forehead, the negative electrode on the low mastoid, and the common electrode on the other mastoid (British Society of Audiology, 2022). The difficulty of routinely setting up a multichannel EEG recording system may challenge the clinical feasibility of measuring cortical responses to continuous speech. Therefore, it was important to explore whether the existing approach to recording and analysing responses to continuous speech could be adapted using single-channel recordings. If successful, continuous speech-evoked potentials could become an additional modality within the AEP system, making them more accessible for use in most audiology clinics. Additionally, using fewer electrodes will simplify clinical measurements, particularly for children. It is important to note that, in the field of measuring cortical responses to continuous speech, the sensitivity of detection at the individual level using single-channel measurements has not been previously explored. As a result, the detectability findings in this thesis cannot be directly compared to those of other studies.

The use of single-channel analysis was explored in the first study (Chapter 3), where pre-existing EEG data from 17 English speakers actively listening to English speech were analysed. The results of this study were promising, as it was possible to achieve 100% detection using both the cross-correlation method after 17 minutes (see Figure 3.3).

The project then investigated whether similar response measurements could be obtained during passive listening, where participants watched a silent movie (Chapter 4). Passive listening while watching something engaging provides a way to maintain young children's attention during cortical response recordings (Ortiz Barajas et al., 2021). In this study (Chapter 4), two different stimuli — natural English speech and speech-modulated noise — were presented to a sample

of 22 English speakers during passive listening. In addition, the study explored the effect of stimulus intelligibility on cortical response detection rates and detection times. A reduction in detection rates was observed with both stimuli compared to the first study, which used the same natural English speech stimulus but with a longer recording. A higher detection rate of 81.8% (18 out of 22) was found when using speech-modulated noise, while the detection of individuals to a natural speech stimulus resulted in a lower rate of 63.6 % (14 out of 22) (see Figure 4.4). Section 7.2.3. provides a discussion of the various types of stimuli used in this thesis.

When comparing the detection rate between the first and second studies, a significant reduction was only detected with the natural speech stimuli results (100% compared to 63.6 % with $p< 0.01$). An explanation for this reduction in detection in the second study is the change to passive attention from the first study, discussed in Section 7.2.4. Another potential cause of the reduced detectability in the second study, which uses natural English speech and speech-modulated noise, could be related to the fact that subjects were focused on a silent movie and reading subtitles. A previous study investigated the use of electrical brain activity signals to enhance machine learning models for semantic language understanding tasks, specifically sentiment analysis and relation detection, during reading comprehension (Hollenstein et al., 2021). The study's findings indicate that EEG signals contain linguistic information related to features of the text measured during a reading task (Hollenstein et al., 2021). Thus, the visual EEG responses from reading our second study might have interfered with auditory response to speech and hence caused noisier recordings compared to the first study.

Finally, the longer recording time in the first study could have contributed to improved detection rates. In the first study, the total recording time was 25 min, but detection reached 100% after 17 min. In studies two and four, the recording time was reduced to 15 min; the experiments were designed based on a more appropriate EEG recording time for clinical settings, particularly for infant testing. Reflecting on the first study, the detection rate after 15 min was slightly reduced from 100% to 94.1% with XCOR. As a result, the shorter recording time provides some explanation for the reduction in detectability in the second study, yet it may combine with other factors, such as attention and the effect of visual processing, leading to the significantly lower detection rates observed in that study.

In the fourth study (Chapter 6), the detectability of the non-language-specific ISTS stimulus was investigated. The detection rate in the attention condition (unaided-70 with attention) after 15 min of recording was 91.6% (22 out of 24), which was not significantly different from the speech results in the first study after 15 min, showing 88.2% (15 out of 17) with XCOR. Since ISTS

showed comparable results to natural speech in terms of detectability, it could be a better option for clinical testing, offering a universally applicable stimulus across various language backgrounds. It is important to highlight that to the best of our knowledge, the ISTS has not previously been used in speech-evoked potentials.

In summary, the results for ISTS detectability in the unaided condition are promising in terms of the applicability of using language-independent tests for future clinical application of cortical responses to continuous speech. However, having fewer than 100% significant responses in normal-hearing individuals is a potential concern in terms of the test's sensitivity to detect the presence of responses in future clinical practice. Therefore, an ideal test would achieve a sensitivity close to 100% in cases where the subject is optimally aided.

7.2.1.1 Aided response detectability:

Before studying the cortical responses to aided continuous speech stimuli in the fourth study (Chapter 6), the effect of HA processing on the stimulus envelope was investigated in the third study (Chapter 5). One objective was to compare the amount of envelope distortion with different stimuli types (natural speech, speech-modulated noise, and ISTS), compression ratios, and input levels. Another main objective was to generate the aided stimuli for the fourth study by recording the output of the HA to different stimuli. The results showed that natural speech and ISTS had similar envelope distortion across different conditions, while speech-modulated noise had significantly lower envelope distortion. A higher compression ratio resulted in significantly higher distortion than low and linear compression ratios, consistent with Souza et al. (2012). Additionally, the study found significantly lower distortion with a 50 dB LAeq input compared to a 70 dB LAeq input. Based on these findings, ISTS stimuli were selected for the subsequent study as they behaved similarly to speech without the distortion seen for modulated noise and, like modulated noise stimuli, were non-language specific. Two aided conditions were selected to investigate how different amounts of envelope distortion affect responses: aided-50 and aided-70, as they had significantly different EDIs, with the EDI being higher in the aided-70 condition. The discussion of these findings was mentioned in Chapter 5 (see Section 5.4).

The fourth study (Chapter 6) investigated the effect of HA processing on stimulus envelope detectability. It should be highlighted that, to the best of our knowledge, this is the first study to investigate how changes in the envelope distortion of aided stimuli affect the resulting neural tracking of the stimulus envelope. In addition, it is the first to measure the detection sensitivity of aided continuous speech cortical responses at an individual level using a single EEG channel.

The cortical responses to ISTS were affected by envelope distortion: McNamar tests showed a significant reduction in aided responses with higher envelope distortion aided-70 (14 out of 24; 58.3%) compared to unaided-70 responses (22 out of 24; 91.7%) (see Figure 6.2 A). The detectability of the aided-50 condition with a lower envelope distortion (20 out of 24; 83.8 %) was not significantly different from the unaided-70 condition (22 out of 24; 91.7%). However, when considering aided testing in the clinic, the compression value, which is related to envelope distortion, may vary widely across subjects. A potential method to improve the detectability of aided responses is to use the stimulus measured from the HA output instead of the original stimulus, although this would need the ability to record the output of the HA in the child's ear, for example using real-ear measurement, which may need additional equipment. This approach showed a significant increase in sensitivity in the aided-70 conditions from 58.3% to 83.3% (see Figure 6.8). However, this rate was still lower than the unaided condition, though the difference was not statistically significant.

In conclusion, cortical responses to aided continuous speech showed less than 100% sensitivity in detecting responses in normal-hearing adults using a single channel after 15 min of recording, with sensitivity further reduced for high levels of HA envelope distortion. For future clinical measurements on infants, it is essential to optimise the stimulus paradigm to ensure responses are reliably detectable in normal-hearing individuals. This is important since some infants do not show detectable responses when a sound is audible. As such, it becomes challenging to determine whether the lack of response with an HA is due to insufficient amplification or a missed neural response. In the current research, a passive listening paradigm using a non-language-specific ISTS stimulation appears promising. However, it is also crucial to account for the distortion of the speech stimulus by an HA, as this may reduce the detectability of brain responses. One approach to consider is investigating methods for measuring the HA output in a child's ear to improve response detection.

7.2.2 Detection time

Test duration is a crucial factor when assessing clinical feasibility. Since measuring cortical responses requires the subject to remain awake and attentive to something such as a screen during the recording, an important question arises: what duration is appropriate for infants? This consideration is vital, as the ultimate goal is to develop a test suitable for this age group. In a CAEP clinical study that included 104 infants, the mean EEG data acquisition time to complete the test for 95% of the sample (94 infants) was a 16 min testing time (Munro et al., 2020). More recently, Visram et al. (2023) conducted aided CAPE testing using two speech-synthetic stimuli

on 103 infants and found a completion rate of 99% with a 14 min to 18 min test duration. Although there are several differences in the testing method between the repeated stimuli used to evoke CAEPs and the continuous speech investigated in this project, the findings of Munro et al. (2020) and Visram et al. (2023) shed light on the issues of test durations for infants in the clinic. Attaheri et al. (2022a), studied neural tracking of the speech envelope measured in 60 infants listening to nursery rhymes whilst watching a video. All infants included in the final analysis were able to attend to a video for about 14 min. However, one limitation of that study was that it did not show only showed significant response at a group level. The detection sensitivity of individual neural tracking was not investigated, leaving it unclear whether the recording time was sufficient at the individual level. Munro et al. (2020) found a 95% detection rate in 16 min, which is broadly similar to the detection rates observed in the first and fourth studies of this project. Based on the time required for CAPE testing to measure significant responses in infants, it is reasonable to suggest that a testing duration of around 14 min to 16 min may be appropriate for measuring cortical responses in this age group. However, speech-evoked responses are typically smaller than standard CAEPs, making them more difficult to detect.

Before the start of this project, the studies investigating the quantity of EEG data required for neural tracking of the speech envelope were limited to a study by Di Liberto and Lalor (2017). They showed that about 30 min of data were required for the subject model to show significant prediction accuracy (detection) in the TRF analysis. The model was built using specific phonemic speech features rather than the overall speech envelope, which might explain the long recording time required. A long recording might be challenging when testing infants. Therefore, using a model that predicts the EEG response based solely on the stimulus envelope or employing a simple cross-correlation approach could reduce the required data quantity, making it a more suitable clinical option. More recently, Mesik and Wojtczak (2022) investigated the effect of EEG data quantity on prediction accuracy by progressively increasing the data range from 3 min to 42 min across 11 steps of data quantities. One interesting finding in their study is that the subject-specific model's performance reached a plateau after about 14 min, when the speech envelope was used as the main detection feature. Although this study progressively increased the amount of data, similar to the approach investigated in the current research, neither the exact detection time needed for each subject nor the maximum detection rate was explored. To the best of our knowledge, the study in this thesis is the first to investigate detection time at the individual level.

When looking into the findings of the current project, in the first study (Chapter 3), the recording time was 25 min, which showed 100% detection after 17 min (see Figure 3.3). This is only a 3 min difference from Mesik and Wojtczak (2022). The longer recording time needed in our study could be explained by the use of single-channel EEG, which requires more data to detect a significant response. To support this point, a preliminary analysis of the data from the first study was conducted using a backward modelling approach. The analysis included 32 EEG channels to compare the backward modelling approach with the two single-channel analyses used in this thesis (forward modelling and XCOR). The results indicated that all subjects exhibited significant responses after 11 min with the backward modelling, compared to 17 min with the single-channel analyses.

The backward modelling analysis approach utilises multichannel data and gives more weight to highly informative channels than to channels with low or no information (Crosse et al., 2016). While it is possible that using a multichannel analysis approach could reduce the required data quantity, this runs contrary to one of the main objectives of the current project, which advocates the use of more straightforward and clinically available single-channel recordings. The increased cost, time, and potential disturbance to the infant when applying multiple electrodes may not be justified by the small reduction in recording time. However, since infants need to remain awake and focused on the screen during the recording of cortical responses, reducing the recording time could be critical. Maintaining the infant's alertness without causing restlessness or excessive movement is essential for obtaining high-quality EEG recordings (Riggins et al., 2007). This, however, becomes more challenging during longer recording sessions.

The recording duration for each condition in the second study (Chapter 4) was reduced to 15 minutes, compared to 25 minutes in the first study (Chapter 3). This change was made for several reasons. First, the detection rate in the first study was 100% after 17 minutes, with a mean detection time of 4.82 minutes. Second, previous literature on testing cortical responses in infants indicated that testing times between 14-18 min had a completion rate of over 95% (Munro et al., 2020, Visram et al., 2023). Finally, the quiet conditions in the second study were part of a separate PhD experiment involving other testing conditions(Deoisres, 2023), with a total experiment duration of about 3 hours, allowing only 15 minutes for testing responses in the two quiet conditions.

As mentioned in Section 7.2.1, the second study showed a significant reduction in detection compared to the first study. As a result, the average detection time for both stimuli used in the second study was higher than the first study: speech XCOR = 11.9 min, speech-modulated

noise XCOR = 9.4 min. The average detection time found in the first study was 4.82 min with XCOR. However, it should be noted that in calculating these averages, subjects showing non-significant responses were allocated a higher detection time of 20 min, which is 5 min higher than the maximum value used in the second study. This may explain the higher average detection time found in the second study compared to the first study, as removing subjects with no detection reduced the mean detection time (speech = 7.36 min and speech-modulated noise = 8.82 min). Additionally, the presence of attention in the first study could have contributed to the shorter detection time (see Section 7.2.4). Therefore, the difference in detection between the first and second studies can be attributed to either the shorter recording time or the passive listening condition.

In the fourth study (Chapter 6), the same recording time was used as in the second study. The average detection time for the unaided-70 condition with passive listening was 6.75 min with XCOR. The average detection time was not significantly different from the first study's mean detection, 4.82 min. This finding adds to the promising use of the ISTS stimulus in measuring cortical responses to continuous speech. For future research, it may be important to investigate detection with a slightly longer recording time to improve detectability, but the appropriate testing time for infants should always be considered.

Another interesting observation in the current project is the high variability between subjects in the quantity of EEG data needed to detect significant cortical responses to speech. While most subject responses can be detected in under 10 min, a small number of subjects require potentially much longer recording times. It would be helpful to understand this large variation in speech response detection times across subjects. Although differences in subject noise levels were explored as a potential cause of the variability in Chapter 6 (see Section 6.3.2), this did not give a clear explanation for the variability seen. These extended detection times for a few subjects may vary between experiments, resulting in different overall detection rates across studies. For future clinical implementation, it can be challenging to determine whether a non-detection is due to subject factors or an issue with HA amplification in difficult detection cases.

7.2.3 Stimulus type

At the initial stage of the project, an intelligible continuous speech stimulus was used to measure neural tracking of the speech envelope. At that stage, the objective was to test the feasibility of single-channel analysis using results from both the TRF and cross-correlation approaches. The intelligibility factor was maintained in line with previous literature that utilised similar analysis methods (Kong et al., 2014, Di Liberto et al., 2015, Di Liberto and Lalor, 2017). In

the second study, the effect of stimulus intelligibility was investigated by comparing detectability and detection time between intelligible English speech and speech-modulated noise in native English speakers with normal hearing. It should also be clear that the attention condition was also changed from the first study to be passive, however, the attention will be discussed in Section 7.2.4.

The speech-modulated noise was developed using the envelope extracted from natural speech, thereby maintaining the envelope. One objective of this study was to assess the feasibility of using a stimulus that mirrors the speech envelope but can be used independently of the subject's language, supporting clinical applicability. The outcomes showed robust detection of cortical responses when using speech-modulated noise, with a high detection rate (see Figure 4.4) and low mean detection times (see Figure 4.6); however, this difference was not statistically significant. One possible explanation is that natural speech contains both amplitude and frequency changes, whereas speech-modulated noise only has amplitude or temporal modulation. Cortical responses vary with changes in temporal information as well as with spectral changes (Okamoto et al., 2012, Vonck et al., 2019). Considering the analysis for studies one, two and four, the amplitude change of the envelope was the only feature correlated to EEG. Therefore, in the case of natural speech, responses related to frequency changes were missed. This might result in lower cortical response detection to natural speech compared to when we have only amplitude-modulated input without the confounding influence of frequency modulation. The modulated noise appears as detectable as natural speech (in fact with slightly higher, but not significantly different, detection rates), which supports the possibility of using speech-modulated noise to objectively assess the perception of speech stimuli.

Studying the impact of intelligibility can be achieved by comparing stimuli with similar acoustic characteristics to speech, such as using noise-vocoded speech (Ding et al., 2014, Kösem et al., 2023), using forging language (Etard and Reichenbach, 2019) presenting speech with different SNR levels (Vanthornhout et al., 2018, Lesenfants et al., 2019), and changing the rate of speech (Verschueren et al., 2022). Some of these studies found that it is possible to predict speech intelligibility from acoustic neural tracking results (Vanthornhout et al., 2018, Lesenfants et al., 2019). Additionally, Ding et al. (2014) found higher cortical entrainment to natural speech than noise-vocoded speech, but only with background noise. In contrast, more recent studies observed no effect of speech intelligibility on acoustic neural tracking (Verschueren et al., 2022, Gillis et al., 2023, Kösem et al., 2023). The discrepancy in the literature about the effect of intelligibility on neural tracking could be explained by the different experimental paradigms used. Specifically, the effect is only evident with the use of added noise to reduce speech

intelligibility. Since changing the SNR is linked to changing speech understanding, it is challenging to determine whether the reduction in neural tracking is primarily caused by diminished speech understanding or by the acoustic change in the signal, reducing the contrast between peaks and troughs in the sound levels used to alter speech comprehension (Verschueren et al., 2022).

The lack of significant differences in the responses between speech and modulated noise could be related to how the brain processes speech and speech-like stimuli. Functional magnetic resonance imaging (fMRI) studies have shown that brain activation in response to intelligible speech occurs in the middle part of the superior temporal sulcus, the ventral pre-central sulcus, and pars opercularis (Binder et al., 2000, Ge et al., 2015). Binder et al. (2000), studied the temporal lobe activation to several types of signals, including unstructured noise, frequency-modulated tone, words, reversed speech and pseudowords using fMRI. The sample included ten normal hearing adult participants. The results showed that direct comparisons between words-pseudowords, words-reversed, and pseudowords-reversed did not reveal any significant differences in area activation across the regions or hemispheres examined.

Although speech-modulated noise showed no different results in detection compared to intelligible speech, the next step was to assess the impact of HA processing on both stimuli. Returning to the primary aim of this thesis, which is to assess the clinical feasibility of using cortical responses to continuous speech for aided speech testing, it was important to examine how HA processing distorts the stimulus envelope and if this then affects the detection of EEG responses. The third study quantified the envelope distortion of the two stimuli used in the second study and a new stimulus, the ISTS. The rationale for testing the ISTS was its combined characteristics of being natural yet language-independent, supporting its potential application to diverse populations worldwide who may not speak the same language. The study concluded that the ISTS showed similar HA-induced distortion as intelligible speech, unlike speech-modulated noise, which exhibited significantly lower HA-induced distortion (see Figure 5.7). One possible explanation is that natural speech contains more variable spectral information. Thus, with more spectral changes, even pure amplification will alter the envelope. In contrast, the specific and controlled acoustic properties of the speech-modulated noise likely made it less susceptible to distortion by HA algorithms. This is because the speech-modulated noise maintains a more predictable signal that has the same amplitude envelope as speech without the complex spectro-temporal changes that create the phonetic content of natural speech.

The fourth study (Chapter 6) used the ISTS stimulus to investigate the feasibility of measuring aided responses, because ISTS behaves similarly to natural speech stimulus during HA

processing, as found in Chapter 5. Although ISTS elicited response detection was affected by a high HA-induced envelope distortion, it still demonstrated similar detectability to natural speech from the first study in both unaided and aided conditions with low distortion.

7.2.4 Attention

In studying cortical or late auditory responses, subject alertness and attention to the stimulus have been shown to enhance neural processing (Fritz et al., 2007, Näätänen et al., 2011). Given that the objective of this thesis was formulated with considerations for infant clinical testing, the effect of attention was investigated. In the first study, only active listening conditions were implemented, which could explain the significantly higher detection of speech in the first studies compared to the second. Studies investigating the effect of attention in neural tracking of continuous speech have shown a non-significant effect in quiet conditions (Kong et al., 2014, Vanthornhout et al., 2019). However, since this thesis explored the feasibility of factors related to infant clinical testing using the single-channel analysis method, the second study examined the possibility of detecting cortical responses to continuous stimuli under passive listening conditions, as controlling infants' attention can be challenging. The detection rate in that study was significantly less than 100% for speech stimuli, but not for speech-modulated noise. Since the study only involved passive listening, the difference between passive and active listening requires further investigation.

One of the objectives of the fourth study was to evaluate the differences between passive and active listening using single-channel analysis, focusing on detectability and recording duration. The main stimulus used in the fourth study was ISTS, which is unintelligible. Consequently, the attention condition in the fourth study involved counting clicks that had been embedded within the ISTS stimulus. The study found no significant effect of attention on either detection rate or detection time. This finding is broadly similar to those of Kong et al. (2014) and Vanthornhout et al. (2019) regarding not having a significant attention effect in quiet conditions. However, previous studies only showed the effect of attention at a group level. To our knowledge, the specific effect of attention on detection rate and detection time at an individual level has not been investigated before.

One possible explanation for the lack of a significant effect between active and passive conditions in the fourth study could be the use of unintelligible ISTS speech, which may make it challenging to maintain attention. Additionally, the task of counting clicks embedded in the stimulus might have been insufficient to maintain attention when the stimulus is not intelligible. Finally, single-channel data analysis might have limited power to detect differences in response

strength due to attention. Overall, the absence of a significant effect of attention is promising for implementing passive listening in infant clinical tests.

7.3 Limitations and Future Work Suggestions

Based on the findings of the studies in this project, future work could focus on two main research areas: (1) optimizing signal processing techniques and (2) investigating other sample characteristics, such as age, since infants and young children may exhibit different cortical responses, potentially leading to varying outcomes in terms of detectability.

7.3.1 Analysis method signal processing

This project concluded that the single-channel analysis approach achieves 100% sensitivity in detecting cortical responses to continuous speech stimuli in adults. However, with aided stimuli, the sensitivity is reduced. One major limitation identified was the limited EEG data available when using single-channel analysis. Given that the primary focus was to assess the clinical feasibility of the methods currently used in the literature, this project concentrated on the clinical and audiological aspects of the measurements and did not extensively explore improvements in signal processing.

Several factors within signal processing could be investigated to enhance detectability. One approach could be using other stimulus features extracted from the speech signal or combinations of features, which may improve the ability to detect responses. Previous studies that utilised other speech features, such as envelope onset, spectrogram and phonetic features, resulted in higher correlation values compared to using only the speech envelope (Di Liberto and Lalor, 2017, Drennan and Lalor, 2019). Although those studies showed higher correlation values, it was not clear whether this would result in higher significance at the individual level, which could be an area for investigation in future studies.

Another suggestion is to improve the method of analysis itself, such as the cross-correlation technique. Previous studies employing the cross-correlation method, such as those by Kong et al. (2014) and Petersen et al. (2017), implemented a slightly different approach. These studies segmented the EEG data into several epochs before cross-correlating them with the stimulus, producing an averaged correlation value. In contrast, the current research applied the XCOR method across the entire length of EEG data. Segmenting the data could potentially enhance the detectability of single-channel analysis, as suggested by Kong et al. (2014), who showed a higher averaged correlation coefficient of 0.07 compared to 0.03 found in the first study of the

current project. Additionally, measuring cortical responses is subject to continuous changes with extended recording time due to variations in the subject's attention, alertness, and other potential noise sources, such as unrelated neural activity. Therefore, studying the correlation across different segments may improve the overall correlation and accuracy of the measurements, thereby enhancing detectability.

A key finding of this project is that ISTS appears superior to other speech stimuli for three reasons: it is highly detectable (> 91%); it can be used universally as it is language-independent; it exhibits an HA processing effect similar to that of natural speech. Moving forward, one possible future approach to improving detectability and reducing recording time with the TRF method would be to develop a generic TRF mode that utilises larger datasets. The concept of generic modelling is based on using data from a group of subjects to train the TRF model and then testing it using data from new subjects (Di Liberto and Lalor, 2017). Compared to subject-specific modelling of TRF, the generic model has shown significant cortical response with lower data quantities (Di Liberto and Lalor, 2017, Mesik and Wojtczak, 2022).

This generic model approach could enhance the clinical feasibility of measuring responses to continuous stimuli. By building a model based on data from a large number of subjects, it could serve as normative data for future measurements. This means that the model, once trained, could be applied to new subjects without the need for extensive individual data collection. In the study by Di Liberto and Lalor (2017), the detection time required for the generic model was 10 min compared to 30 min for the subject-specific model. When testing infants, the development of the generic model could significantly reduce the burden on the patients, their parents, and the clinicians. In addition, if this method proves to be successful in terms of detection sensitivity in normal hearing subjects, it could be implemented for real-time detection without the need to train the model for each subject.

It is important to note that, in clinical work, continuous speech-evoked responses may generally be smaller than standard AEPs, making them more difficult to detect. While state-of-the-art methods and multi-channel analysis might overcome this challenge, it has yet to be demonstrated in clinical applications, and multi-channel recordings remain a challenge for clinical use.

7.3.2 Research participants

The scope of the current research was limited to testing adults, with normal hearing, as in the second and fourth studies (Chapter 4 and Chapter 6). At this initial stage of assessing the

clinical feasibility of measuring cortical responses to continuous speech, focusing on this sample group was a reasonable first step for several reasons. Normal-hearing adults provide a controlled baseline without confounding variability caused by hearing or related cognitive impairments. Also, as one of the project objectives is to study the difference between active and passive listening conditions, it is more feasible to control attention in adults compared to infants and young children. Finally, infant recruitment is prone to significant ethical challenges.

The first study showed a 100% detectability after 17 min. Therefore, a possible future step could involve testing normal-hearing infants using a similar approach to assess detectability with a single channel. There is a possibility of observing different outcomes with infants, as shown in Attaheri et al. (2022b), who found differences in neural envelope tracking between adults and infants. This difference was based on the Power Spectral Density (PSD) analysis used by Attaheri et al. (2022b). The PSD explains how the power of a signal or time series is distributed across different frequencies (Unde and Shriram, 2014). Attaheri et al. (2022b) found that neural tracking in infants was more closely tied to the acoustic characteristics of the stimulus compared to adults. In infants, the PSD peaks occurred at frequencies (2.20 Hz and 4.37 Hz) that corresponded directly to the modulation spectrum of nursery rhymes. In contrast, adults exhibited PSD peaks at lower frequencies (1.25 Hz), indicating a less stimulus-driven response. This suggests that while infants' brain activity is more influenced by the rhythmic properties of the sung speech, adults may engage additional cognitive processes, possibly related to comprehension or linguistic experience. Moreover, previous studies in infants have shown significant reconstruction correlation using the backward modelling at the group level (Jessen et al., 2019, Attaheri et al., 2022a, Attaheri et al., 2022b). Future studies could use the single-channel approach and ISTS stimulus to study individual detectability in normal-hearing infants. Achieving this will test the validity of using this test for clinical assessment of cortical responses to speech stimulus through HAs.

Subject variability is another essential aspect to investigate further, especially concerning detection time. This variability might be even larger in infants, as CAEP responses showed significant variability among infants, as reported by (Visram et al., 2023). The findings in this thesis highlighted a high variability in detection time, ranging from 1 min to 17 min. In addition, it seems that a few subjects have responses that are very difficult to measure. Whether this is due to subject noise, or small brain responses is still unclear. Investigating the repeatability of these responses at the individual level could reveal whether the testing parameters are consistent within subjects or influenced by other factors, such as individual alertness or the quality of the recording.

7.4 Conclusion

In summary, this thesis explored several key factors related to the clinical feasibility of using cortical responses to continuous speech stimuli, with a particular focus on optimising the test protocol, assessing the feasibility of single-channel analysis, and the future aim of testing infants. This work may be necessary before an effective determination of whether it will be the 'optimal' approach for clinical purposes — a conclusion that would require a larger clinical study involving hearing-impaired infants.

While the use of ISTS stimuli and passive listening has shown promise, the less-than-optimal detection rate, especially with aided stimulation, and the variability in detection time are areas that require further investigation. Extending this research to optimise detection sensitivity through enhanced signal analysis methods and exploring detection in normal-hearing infants could be the next steps toward developing an objective clinical tool to assess speech perception through hearing aids.

This area of research must continue to bridge the gap between laboratory studies and actual clinical applications. Addressing these challenges is crucial for future advancements in auditory diagnostics and evaluating interventions for hearing loss.

Appendix A Code of Analysis TRF

```
load EEG_data.mat % preprocessed EEG
load Stimulus_envelope.mat %stimulus envelope

ISTS_ENV=ISTS_ENV/std(ISTS_ENV); %normalise envelope
i=30; % channel to use
data1=data(i,:);
data1=data1/std(data1); %normalise
data1=double(data1); % convert to double
data1=data1';
%data1=fliplr(data1')'; % decorrelate by flipping data!

count=1;
for i=1:100
    sigparts(i)=0; % count significant decoder responses in sigparts
end

% % Initialize variables
% sigparts = zeros(1, 100);    % To count significant decoder responses

% envnew = [];
% datanew = [];

envnew=0;
datanew=0;
% artefact rejection
reject=5;
countrej=0;
% countgood = []; % Accepted sweeps
% countall = [];  % All sweeps
countgood=0; % keep track of the accepted sweeps
countall=0; % all sweeps

for i=1:fix(length(ISTS_ENV)/100)
    temp=data1(((i-1)*100+1):(i*100));
    tempenv=ISTS_ENV(((i-1)*100+1):(i*100));
    if max(temp)>reject
        countrej=countrej+1;
    elseif min(temp)<-reject
        countrej=countrej+1;
    else
    countgood=[countgood i]; % good sweep numbers
    datanew=[datanew' temp']'; % add good data
    envnew=[envnew' tempenv']'; % add corresponding good envelope
    end
    countall=[countall i];
end
% Replace data1 and ISTS_ENV with artifact-free data
data1=datanew;
ISTS_ENV=envnew;

% Main loop to process data
for T=7680:7680:length(data1') % loop time in min
    A=ISTS_ENV(1:T); % take first segment of envelope
    B=data1(1:T); % take first segment of data
   % Bootstrapping process
    for n=1:500 % increase to 500 rotations
```

```
        x=(fix(rand*(T-2000)+1999));
        Env=[A(x+1:length(A))' A(1:x)']'; % rotate envelope

        % Train model
        [w,t] = mTRFtrain(Env,B,128,1,-150,450,1e4);
        W(n,:)=w; % store random rotations

        % Filter the envelope and calculate correlation
        output=filter(w,1,Env); % filter the envelope with the filter from
mTRFTtrain
        [R,P]=corrcoef(output,B); % calculate correlation
        %CORBOOT(n)=R(2,1);
        R=xcorr(output,B,'coeff');
        middlepoint=fix(length(R)/2); % find middle of R
        CORBOOT(n)=max(abs(R(middlepoint-1000:middlepoint+1000))); % Max XCOR
        PowerBoot(n)=var(w(21:69)); % Variance for specific time window
    end
    for i=1:79
        V=sort(W(:,i));
%         lower5(i)=V(5);
%         upper5(i)=V(495);
    end
    % now find max and min values for the 500 bootstraps (over 79 samples)
    % Sort max and min points from bootstraps
    maxpoints=max(W');
    minpoints=min(W');
    maxpointssort = sort(maxpoints);
    minpointssort = sort(minpoints);
    lowerT = minpointssort(13); % Lower 2.5% of min points
    upperT = maxpointssort(488); % Upper 2.5% of max points

    % Train on the actual data
    [w,t] = mTRFtrain(A,B,128,1,-150,450,1e4);

   figure;
   plot(t,ones(79)*lowerT);hold on;plot(t,ones(79)*upperT);
   plot(t,w,'k','linewidth',4);
   xlabel('Time lag (ms)');
   ylabel('Amplitude (a.u.)');
   xlim([-50, 350]);

    output=filter(w,1,A); % filter the envelope with the filter from mTRFTtrain
    R=xcorr(output,B,'coeff');
    middlepoint=fix(length(R)/2); % find middle of R
    COR(count)=max(abs(R(middlepoint-1000:middlepoint+1000))); % Max XCOR
    Power(count)=var(w(21:69)); % include a power measurement

   % Store bootstrapped bounds
    lowers(count,:)=lowerT; % use lower 5% of min values
    uppers(count,:)=upperT; % use upper 5% of max values
    mtrf(count,:)=w;

    % Calculate significant parts
    for i=21:69 % run from 0 to 375 ms - avoid end effects
        if w(i)>upperT
            sigparts(count)=sigparts(count)+w(i); % add significant positive
values
        end
        if w(i)<lowerT % use lower 5% of max values
```

```
            sigparts(count)=sigparts(count)-w(i); % add significant positive
values
        end
    end

    % Calculate p-value for correlation
    countB = sum(COR(count) > CORBOOT);
    sigCOR(count) = 1 - (countB / 500); % P-value

     % Calculate p-value for power
    countC = sum(Power(count) > PowerBoot);
    sigPower(count) = 1 - (countC / 500); % P-value

    % Calculate significance of max or min in the time window - 'sigPeak'
    Maxpoint=max(w(21:69)); % find max in time window
    Minpoint=min(w(21:69)); % find min point
    if Maxpoint > -Minpoint
        CountD = sum(Maxpoint > maxpointssort);
    else
        CountD = sum(Minpoint < minpointssort);
    end
    sigPeak(count) = 1 - (CountD / 500); % P-value

    % Track time in minutes
    temp6=find(countall==countgood(fix(count*7680/100)));
    timepoint(count)=temp6*100/(128*60); % time in mins of data

    % Save bootstrap mTRFs
    BootstrapMTRFs(count, :, :) = W;

    % Increment counter
    count = count + 1;
end

% Save results
save('TRFClean_ResultISTS.mat', 'COR', 'sigparts', 'uppers', 'lowers', 'mtrf',
't', 'BootstrapMTRFs', 'sigCOR', 'sigPower', 'sigPeak', 'timepoint');
```

Appendix B Code of Analysis XCOR

```
load EEG_data.mat % preprocessed EEG
load Stimulus_envelope.mat %stimulus envelope

ISTS_ENV=ISTS_ENV/std(ISTS_ENV); %normalise envelope
i=30; % channel to use
data1=data(i,:);
data1=data1/std(data1); %normalise
data1=double(data1); % convert to double
data1=data1';
%data1=fliplr(data1')'; % decorrelate by flipping data!

envnew=0;
datanew=0;
% add artefact rejection
reject=5; % This might need changing
count=0;
countgood=0; % keep track of the accepted sweeps
countall=0; % all sweeps

% ISTS_ENV = transpose(ISTS_ENV);
for i=1:fix(length(data1)/100)
    temp=data1(((i-1)*100+1):(i*100));
    tempenv=ISTS_ENV(((i-1)*100+1):(i*100));
    if max(temp)>reject
        count=count+1;
    elseif min(temp)<-reject
        count=count+1;
    else
    countgood=[countgood i]; % good sweep numbers
    datanew=[datanew' temp']'; % add good data
    envnew=[envnew' tempenv']'; % add corresponding good envelope
    end
    countall=[countall i];
end

counta=1;
sigparts(1:100)=0; % parameter to do with 'sigificant' correlations
maxparts(1:100)=0; % maximum correlation from -0.5 to +0.5 s
endpoints = 67; % Skip initial/final data to avoid delay issues
delays = 64; % Delays to measure over
pValues = []; % Store p-values for each minute
% Main loop processing the data
for L=(7680+(endpoints*2)):7680:length(datanew) % try to loop length of data
    ENV=envnew(endpoints:(L-endpoints)); % remove ends
    result = zeros(1, 2 * delays + 1); % Initialise result
    for n=-delays:delays % delays to measure over
        data2=datanew(endpoints+n:(L-endpoints)+n); % shifted version of data
        C=corrcoef(ENV,data2); % work out correlation
        result(n+delays+1)=C(1,2); % start results with index 1
    end
    % Bootstrap for significance testing
    resultboot = zeros(500, 2 * delays + 1);
    upper = zeros(500, 1);
    lower = zeros(500, 1);
    for nnn=1:500 % boostraps
    x=(fix(rand*(length(datanew)-2000)+1999)); % add a random shift at least 2000
```

```
    databoot=[datanew(x+1:length(datanew))' datanew(1:x)']'; % rotate data
randomly
        for n=-delays:delays % delays to measure over
        data2=databoot(endpoints+n:(L-endpoints)+n); % shifted version of data
        C=corrcoef(ENV,data2); % work out correlation
        resultboot(nnn,n+delays+1)=C(1,2); % start results with index 1
        end
        temp=resultboot(nnn,(delays):(delays+48)); % choose important section
        upper(nnn)=max(temp); % max value of bootstrap correlation in RoI
        lower(nnn)=min(temp); % min value of bootstrap correlation in RoI
    end

 % Calculate p-value for maximum absolute XCOR
    absMaxOriginal = max(abs(result(64:112)));
    absMaxBootstrap = max(abs(resultboot(:, 64:112)), [], 2);
    greaterCount = sum(absMaxBootstrap >= absMaxOriginal);
    pValues(counta) = greaterCount / 500; % Calculate p-value

% Sort bootstrap results and define 2.5% thresholds
uppersort=sort(upper);
lowersort=sort(lower);
upperN(1:length(result))=uppersort(488); % upper 2.5% of max values
lowerN(1:length(result))=lowersort(13); % bottom lower 2.5% of min values

% Plot result
figure;plot(result); xlim([(delays-64) (delays+64)])
hold on;plot(lowerN,'g');plot(upperN,'r')

% save results
Allresults(counta,:)=result;
Alluppers(counta,:)=upperN;
Alllowers(counta,:)=lowerN;
resultboot_all(counta,:,:) = resultboot;

% Find significant parts
for d=(delays):(delays+48) % 19/12 changed to count over +ve time 48 samples -
same as decoder
    if result(d)>uppersort(488) % upper 2.5% of max values
        sigparts(counta)=sigparts(counta)+result(d); % look over 1s. Add
significant responses
    end
    if result(d)<lowersort(13) % bottom lower 2.5% of min values
        sigparts(counta)=sigparts(counta)-result(d);
    end
end

temp6=find(countall==countgood(fix(counta*7680/100)));
timepoint(counta)=temp6*100/(128*60); % time in mins of data
   % timepoint(counta)=find(countall=countgood(L));
counta=counta+1 % index for sigparts

% Plot significant parts vs. minutes forward
figure;
plot(sigparts(1:counta-1));
title('Sum of significant parts vs. minutes forward');
forwardresult = sigparts(1:counta-1);

save Clean_Result_1 'forwardresult' 'Allresults' 'Alluppers' 'Alllowers'
'timepoint' 'pValues'
end
```

List of References

AHISSAR, E., NAGARAJAN, S., AHISSAR, M., PROTOPAPAS, A., MAHNCKE, H. & MERZENICH, M. M. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences USA,* 98**,** 13367-72.

AIKEN, S. J. & PICTON, T. W. 2006. Envelope following responses to natural vowels. *Audiology and Neurotology,* 11**,** 213-32.

AIKEN, S. J. & PICTON, T. W. 2008a. Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research,* 245**,** 35-47.

AIKEN, S. J. & PICTON, T. W. 2008b. Human cortical responses to the speech envelope. *Ear and Hearing,* 29**,** 139-57.

ALICKOVIC, E., LUNNER, T., WENDT, D., FIEDLER, L., HIETKAMP, R., NG, E. H. N. & GRAVERSEN, C. 2020. Neural Representation Enhanced for Speech and Reduced for Background Noise With a Hearing Aid Noise Reduction Scheme During a Selective Attention Task. *Frontiers in Neuroscience,* 14**,** 846.

ALICKOVIC, E., NG, E. H. N., FIEDLER, L., SANTURETTE, S., INNES-BROWN, H. & GRAVERSEN, C. 2021. Effects of Hearing Aid Noise Reduction on Early and Late Cortical Representations of Competing Talkers in Noise. *Frontiers in Neuroscience,* 15**,** 1-20.

ALJARBOA, G. S., BELL, S. L. & SIMPSON, D. M. 2023. Detecting cortical responses to continuous running speech using EEG data from only one channel. *International Journal of Audiology,* 62**,** 199-208.

AMERICAN SPEECH LANGUAGE HEARING ASSOCIATION (ASHA). 2022. *Effects of Hearing Loss on Development* [Online]. American Speech-Language-Hearing Association. Available: https://www.asha.org/public/hearing/effects-of-hearing-loss-on-development/ [Accessed 9 Sep 2022].

ATTAHERI, A., CHOISDEALBHA Á, N., DI LIBERTO, G. M., ROCHA, S., BRUSINI, P., MEAD, N., OLAWOLE-SCOTT, H., BOUTRIS, P., GIBBON, S., WILLIAMS, I., GREY, C., FLANAGAN, S. & GOSWAMI, U. 2022a. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage,* 247**,** 118698.

ATTAHERI, A., PANAYIOTOU, D., PHILLIPS, A., NÍ CHOISDEALBHA, Á., DI LIBERTO, G. M., ROCHA, S., BRUSINI, P., MEAD, N., FLANAGAN, S., OLAWOLE-SCOTT, H. & GOSWAMI, U. 2022b. Cortical Tracking of Sung Speech in Adults vs Infants: A Developmental Analysis. *Frontiers in Neuroscience,* 16**,** 842447.

ATTIAS, J., NAGERIS, B., RALPH, J., VAJDA, J. & RAPPAPORT, Z. H. 2008. Hearing preservation using combined monitoring of extra-tympanic electrocochleography and auditory brainstem responses during acoustic neuroma surgery. *International Journal of Audiology,* 47**,** 178-84.

BARAJAS, J. J. 1985. Brainstem response audiometry as subjective and objective test for neurological diagnosis. *Scandinavian Audiology,* 14**,** 57-62.

BENCH, J., KOWAL, A. & BAMFORD, J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology,* 13**,** 108-12.

BILLINGS, C. J., TREMBLAY, K. L., SOUZA, P. E. & BINNS, M. A. 2007. Effects of Hearing Aid Amplification and Stimulus Intensity on Cortical Auditory Evoked Potentials. *Audiology and Neurotology,* 12**,** 234-246.

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., BELLGOWAN, P. S., SPRINGER, J. A., KAUFMAN, J. N. & POSSING, E. T. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex,* 10**,** 512-28.

BINKHAMIS, G., ELIA FORTE, A., REICHENBACH, T., O'DRISCOLL, M. & KLUK, K. 2019. Speech Auditory Brainstem Responses in Adult Hearing Aid Users: Effects of Aiding and Background Noise, and Prediction of Behavioral Measures. *Trends in Hearing,* 23**,** 1-20.

BLINOWSKA, K. & DURKA, P. 2006. Electroencephalography (EEG). *In:* AKAY, M. (ed.) *Wiley Encyclopedia of Biomedical Engineering*.

BOERSMA, P. & VAN HEUVEN, V. 2001. Speak and unSpeak with PRAAT. *Glot International,* 5 341-347.

BOOTHROYD, A., HANIN, L. & HNATH, T. 1985. A sentence test of speech perception: reliability, set equivalence, and short term learning. The City University of New York (CUNY) Graduate Center.

BRIGHT, K., GREELEY, C., EICHWALD, J., LOVELAND, C. & TANNER, G. 2011. American Academy of Audiology childhood hearing screening guidelines. *Reston, VA: American Academy of Audiology Task Force*.

BRITISH SOCIETY OF AUDIOLOGY. 2018. *Practice Guidance: Assessment of speech understanding in noise in adults with hearing difficulties* [Online]. BRITISH SOCIETY OF AUDIOLOGY. Available: https://www.thebsa.org.uk/wp-content/uploads/2019/04/OD104-80-BSA-Practice-Guidance-Speech-in-Noise-FINAL.Feb-2019.pdf [Accessed 22 July 2020].

BRITISH SOCIETY OF AUDIOLOGY. 2022. *Cortical Auditory Evoked Potential (CAEP) Testing* [Online]. Available: https://www.thebsa.org.uk/wp-content/uploads/2023/10/OD104-40-BSA-Recommended-Procedure-CAEP.pdf [Accessed 29 April 2024].

BROWN, E., KLEIN, A. J. & SNYDEET, K. A. 1990. Hearing-aid-processed tone pips: electroacoustic and ABR characteristics. *Journal of the American Academy of Audiology* 10 190-197.

CARTER, L., GOLDING, M., DILLON, H. & SEYMOUR, J. 2010. The detection of infant cortical auditory evoked potentials (CAEPs) using statistical and visual detection techniques. *Journal of the American Academy of Audiology* 21**,** 347-56.

CHALAS, N., DAUBE, C., KLUGER, D. S., ABBASI, O., NITSCH, R. & GROSS, J. 2022. Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex. *NeuroImage,* 258**,** 1-11.

CHANDRASEKARAN, B. & KRAUS, N. 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology,* 47**,** 236-246.

CHESNAYE, M. A., BELL, S. L., HARTE, J. M., SIMONSEN, L. B., VISRAM, A. S., STONE, M. A., MUNRO, K. J. & SIMPSON, D. M. 2023. Modified T(2) Statistics for Improved Detection of

Aided Cortical Auditory Evoked Potentials in Hearing-Impaired Infants. *Trends in Hearing,* 27**,** 1-17.

CHESNAYE, M. A., BELL, S. L., HARTE, J. M. & SIMPSON, D. M. 2018. Objective measures for detecting the auditory brainstem response: comparisons of specificity, sensitivity and detection time. *International Journal of Audiology,* 57**,** 468-478.

CHING, T. Y. 2015. Is Early Intervention Effective in Improving Spoken Language Outcomes of Children With Congenital Hearing Loss? *American Journal of Audiology,* 24**,** 345-8.

CHING, T. Y. & HILL, M. 2007. The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale: normative data. *Journal of the American Academy of Audiology* 18**,** 220-35.

CHINNARAJ, G., SHETTY, H. N. & PALANIYAPPAN, V. 2021. EFFECT OF DIGITAL NOISE REDUCTION AND DIRECTIONALITY ALGORITHMS IN HEARING AIDS ON TEMPORAL ENVELOPE DISTORTION AND SPEECH RECOGNITION. *Journal of Hearing Science,* 11**,** 19-29.

CHOI, J. M., PURCELL, D. W., COYNE, J. A. & AIKEN, S. J. 2013. Envelope following responses elicited by English sentences. *Ear and Hearing,* 34**,** 637-650.

COX, R. M. & ALEXANDER, G. C. 1995. The abbreviated profile of hearing aid benefit. *Ear and Hearing,* 16**,** 176-86.

CROSSE, M. J., DI LIBERTO, G. M., BEDNAR, A. & LALOR, E. C. 2016. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience,* 10**,** 1-14.

DECRUY, L., VANTHORNHOUT, J. & FRANCART, T. 2020. Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research,* 393**,** 1-13.

DEOISRES, S. 2023. *Auditory cortical responses as an objective predictor of speech-in-noise performance*. PhD, University of Southampton

DEOISRES, S., LU, Y., VANHEUSDEN, F. J., BELL, S. L. & SIMPSON, D. M. 2023. Continuous speech with pauses inserted between words increases cortical tracking of speech envelope. *PLoS One,* 18**,** e0289288.

DI LIBERTO, G. M. & LALOR, E. C. 2017. Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. *Hearing Research,* 348**,** 70-77.

DI LIBERTO, G. M., O'SULLIVAN, J. A. & LALOR, E. C. 2015. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology,* 25**,** 2457-65.

DILLON, H. 2005. So, baby, how does it sound? Cortical assessment of infants with hearing aids. *The Hearing Journal,* 58**,** 12-17.

DILLON, H. 2012. *Hearing aids / Harvey Dillon*, Boomerang Press ; Thieme.

DIMITRIJEVIC, A. & CONE-WESSON, B. K. 2015. Auditory Steady-State Response. *In:* KATZ, J., CHASIN, M., ENGLISH, K., HOOD, L. & TILLERY, K. (eds.) *Handbook of Clinical Audiology*.

DIMITRIJEVIC, A., JOHN, M. S. & PICTON, T. W. 2004. Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear and Hearing,* 25**,** 68-84.

DING, N., CHATTERJEE, M. & SIMON, J. Z. 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage,* 88**,** 41-6.

DING, N. & SIMON, J. Z. 2012a. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences USA,* 109**,** 11854-9.

DING, N. & SIMON, J. Z. 2012b. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology,* 107**,** 78-89.

DING, N. & SIMON, J. Z. 2013. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *The Journal of Neuroscience,* 33**,** 5728-5735.

DRENNAN, D. P. & LALOR, E. C. 2019. Cortical Tracking of Complex Sound Envelopes: Modeling the Changes in Response with Intensity. *eNeuro,* 6.

DRULLMAN, R., FESTEN, J. M. & PLOMP, R. 1994. Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America,* 95**,** 1053-1064.

EASWAR, V., BEAMISH, L., AIKEN, S., CHOI, J. M., SCOLLIE, S. & PURCELL, D. 2015a. Sensitivity of envelope following responses to vowel polarity. *Hearing Research,* 320**,** 38-50.

EASWAR, V., PURCELL, D. W., AIKEN, S. J., PARSA, V. & SCOLLIE, S. D. 2015b. Effect of Stimulus Level and Bandwidth on Speech-Evoked Envelope Following Responses in Adults With Normal Hearing. *Ear and Hearing,* 36**,** 619-34.

EASWAR, V., PURCELL, D. W., AIKEN, S. J., PARSA, V. & SCOLLIE, S. D. 2015c. Evaluation of Speech-Evoked Envelope Following Responses as an Objective Aided Outcome Measure: Effect of Stimulus Level, Bandwidth, and Amplification in Adults With Hearing Loss. *Ear and Hearing,* 36**,** 635-52.

EASWAR, V., PURCELL, D. W. & SCOLLIE, S. D. 2012. Electroacoustic Comparison of Hearing Aid Output of Phonemes in Running Speech versus Isolation: Implications for Aided Cortical Auditory Evoked Potentials Testing. *International Journal of Otolaryngology,* 2012**,** 1-10.

EASWAR, V., SCOLLIE, S., AIKEN, S. & PURCELL, D. 2020. Test-Retest Variability in the Characteristics of Envelope Following Responses Evoked by Speech Stimuli. *Ear and Hearing,* 41**,** 150-164.

EISENBERG, L., JOHNSON, K. C. & MARTINEZ, A. S. 2005. *Clinical Assessment of Speech Perception for Infants and Toddlers* [Online]. Available: https://www.audiologyonline.com/articles/clinical-assessment-speech-perception-for-1016 [Accessed 7 Sep 2020].

EISENBERG, L. S., JOHNSON, K. C. & MARTINEZ, A. S. 2010. Measuring Auditory Performance of Pediatric Cochlear Implant Users : What Can Be Learned for Children who Use Hearing Instruments ? *In:* SEEWALD, R., ed. A Sound Foundation Through Early Amplification, 2010.

ER, E. R. 2011. *Quick Speech-in-Noise Test (Version 1.3) - User manual* [Online]. Etymotic Research. Available:

https://www.etymotic.com/downloads/dl/file/id/259/product/159/quicksin_user_manual.pdf [Accessed 21 Sep 2020].

ERBER, N. 1976. The use of audio tape-cards in auditory training for hearing-impaired children. *The Volta review,* 78**,** 209-218.

ETARD, O., KEGLER, M., BRAIMAN, C., FORTE, A. E. & REICHENBACH, T. 2019. Decoding of selective attention to continuous speech from the human auditory brainstem response. *NeuroImage,* 200**,** 1-11.

ETARD, O. & REICHENBACH, T. 2019. Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. *The Journal of Neuroscience,* 39**,** 5750 -5759.

FERNALD, A., TAESCHNER, T., DUNN, J., PAPOUSEK, M., DE BOYSSON-BARDIES, B. & FUKUI, I. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language,* 16**,** 477-501.

FORTE, A. E., ETARD, O. & REICHENBACH, T. 2017. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife,* 6.

FORTUNE, T. W., WOODRUFF, B. D. & PREVES, D. A. 1994. A new technique for quantifying temporal envelope contrasts. *Ear and Hearing,* 15**,** 93-9.

FRITZ, J. B., ELHILALI, M., DAVID, S. V. & SHAMMA, S. A. 2007. Auditory attention—focusing the searchlight on sound. *Current Opinion in Neurobiology,* 17**,** 437-455.

GALBRAITH, G. C., ARBAGEY, P. W., BRANSKI, R., COMERCI, N. & RECTOR, P. M. 1995. Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport,* 6**,** 2363-7.

GARDNER-BERRY, K., CHANG, H., CHING, T. Y. & HOU, S. 2016. Detection rates of cortical auditory evoked potentials at different sensation levels in infants with sensory/neural hearing loss and auditory neuropathy spectrum disorder. *Seminars in Hearing,* 37**,** 53.

GARNHAM, J., COPE, Y., DURST, C., MCCORMICK, B. & MASON, S. M. 2000. Assessment of aided ABR thresholds before cochlear implantation. *British Journal of Audiology,* 34**,** 267-78.

GE, J., PENG, G., LYU, B., WANG, Y., ZHUO, Y., NIU, Z., TAN, L., LEFF, A. & GAO, J.-H. 2015. Cross-language differences in the brain network subserving intelligible speech. *Proceedings of the National Academy of Sciences USA,* 112.

GEERS, A. E. & MOOG, J. S. 2012. Early speech perception. *CID early speech perception*. St. Louis, Missouri: Central Institute for the Deaf.

GEETHA, C. & MANJULA, P. 2014. Effect of Compression, Digital Noise Reduction and Directionality on Envelope Difference Index, Log-Likelihood Ratio and Perceived Quality. *Audiology Research,* 4**,** 110.

GIBSON, W. P. 2017. The Clinical Uses of Electrocochleography. *Frontiers in Neuroscience,* 11**,** 274-274.

GILLIS, M., DECRUY, L., VANTHORNHOUT, J. & FRANCART, T. 2022a. Hearing loss is associated with delayed neural responses to continuous speech. *European Journal of Neuroscience,* 55**,** 1671-1690.

GILLIS, M., VAN CANNEYT, J., FRANCART, T. & VANTHORNHOUT, J. 2022b. Neural tracking as a diagnostic tool to assess the auditory pathway. *Hearing Research,* 426**,** 1-14.

GILLIS, M., VANTHORNHOUT, J. & FRANCART, T. 2023. Heard or Understood? Neural Tracking of Language Features in a Comprehensible Story, an Incomprehensible Story and a Word List. *eNeuro,* 10**,** ENEURO.0075-23.2023.

GINSBERG, H., SINGH, R., BHARADWAJ, H. M. & HEINZ, M. G. 2023. A multi-channel EEG mini-cap can improve reliability for recording auditory brainstem responses in chinchillas. *Journal of Neuroscience Methods,* 398**,** 109954.

GOLDING, M., DILLON, H., SEYMOUR, J. & CARTER, L. 2009. The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. *International Journal of Audiology,* 48**,** 833-42.

GOLDING, M., PEARCE, W., SEYMOUR, J., COOPER, A., CHING, T. & DILLON, H. 2007. The relationship between obligatory cortical auditory evoked potentials (CAEPs) and functional measures in young infants. *Journal of the American Academy of Audiology* 18**,** 117-25.

GORGA, M. P., BEAUCHAINE, K. A. & REILAND, J. K. 1987. Comparison of onset and steady-state responses of hearing aids: implications for use of the auditory brainstem response in the selection of hearing aids. *Journal of Speech, Language, and Hearing Research,* 30**,** 130-6.

GRILL-SPECTOR, K., HENSON, R. & MARTIN, A. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences,* 10**,** 14-23.

HALL, J. W. J. W. 2015. *eHandbook of Auditory Evoked Responses*.

HAMILTON, L. S., EDWARDS, E. & CHANG, E. F. 2018. A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Current Biology,* 28**,** 1860-1871.

HANSEN, C. H. 2001. Fundamentals of acoustics. *Occupational Exposure to Noise: Evaluation, Prevention and Control. World Health Organization,* 1**,** 23-52.

HOLLENSTEIN, N., RENGGLI, C., GLAUS, B., BARRETT, M., TROENDLE, M., LANGER, N. & ZHANG, C. 2021. Decoding EEG brain activity for multi-modal natural language processing. *Frontiers in Human Neuroscience***,** 378.

HOLUBE, I., FREDELAKE, S., VLAMING, M. & KOLLMEIER, B. 2010. Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology,* 49**,** 891-903.

HOWARD, M. F. & POEPPEL, D. 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology,* 104**,** 2500-11.

INTERACOUSTICS. 2022. *Electrode tips and tricks for Eclipse ABR testing* [Online]. Available: https://www.interacoustics.com/abr-equipment/eclipse/support/electrode-tips-and-tricks [Accessed 14 Jan 2024].

JENSTAD, L. M. & SOUZA, P. E. 2005. Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language, and Hearing Research,* 48**,** 651-67.

JESSEN, S., FIEDLER, L., MÜNTE, T. F. & OBLESER, J. 2019. Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *NeuroImage,* 202**,** 116060.

KALASHNIKOVA, M., PETER, V., DI LIBERTO, G. M., LALOR, E. C. & BURNHAM, D. 2018. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports,* 8**,** 13745.

KILNER, J. M. 2013. Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology,* 124**,** 2062-3.

KONG, Y. Y., MULLANGI, A. & DING, N. 2014. Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research,* 316**,** 73-81.

KÖSEM, A., DAI, B., MCQUEEN, J., M. & HAGOORT, P. 2023. Neural tracking of speech envelope does not unequivocally reflect intelligibility. *NeuroImage,* 272**,** 120040.

KOWALEWSKI, B., ZAAR, J., FERECZKOWSKI, M., MACDONALD, E. N., STRELCYK, O., MAY, T. & DAU, T. 2018. Effects of Slow- and Fast-Acting Compression on Hearing-Impaired Listeners' Consonant-Vowel Identification in Interrupted Noise. *Trends in Hearing,* 22**,** 2331216518800870.

KRAUS, N., ANDERSON, A. J. & WHITE-SCHWOCH, T. 2017. *The Frequency-Following Response: A Window into Human Communication*, Springer Nature.

LALOR, E. C. & FOXE, J. J. 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience,* 31**,** 189-93.

LALOR, E. C., POWER, A. J., REILLY, R. B. & FOXE, J. J. 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology,* 102**,** 349-59.

LAUGESEN, S., RIECK, J. E., ELBERLING, C., DAU, T. & HARTE, J. M. 2018. On the Cost of Introducing Speech-Like Properties to a Stimulus for Auditory Steady-State Response Measurements. *Trends in Hearing,* 22**,** 2331216518789302.

LESENFANTS, D., VANTHORNHOUT, J., VERSCHUEREN, E., DECRUY, L. & FRANCART, T. 2019. Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. *Hearing Research,* 380**,** 1-9.

LIGHTFOOT, G. 2016. Summary of the N1-P2 Cortical Auditory Evoked Potential to Estimate the Auditory Threshold in Adults. *Seminars in Hearing,* 37**,** 1-8.

LIGHTFOOT, G. & KENNEDY, V. 2006. Cortical electric response audiometry hearing threshold estimation: accuracy, speed, and the effects of stimulus presentation features. *Ear Hear,* 27**,** 443-56.

LUCK, S. J. 2014. *An Introduction to the Event-Related Potential Technique,* Cambridge, UNITED STATES, MIT Press.

LV, J., SIMPSON, D. M. & BELL, S. L. 2007. Objective detection of evoked potentials using a bootstrap technique. *Medical Engineering and Physics,* 29**,** 191-198.

MCARDLE, R. & HNATH-CHISOLM, T. 2015. Speech Audiometry *In:* KATZ, J., CHASIN, M., ENGLISH, K., HOOD, L. & TILLERY, K. (eds.) *Handbook of Clinical Audiology*.

MEINZEN-DERR, J., WILEY, S., GROVE, W., ALTAYE, M., GAFFNEY, M., SATTERFIELD-NASH, A., FOLGER, A. T., PEACOCK, G. & BOYLE, C. 2020. Kindergarten Readiness in Children Who Are Deaf or Hard of Hearing Who Received Early Intervention. *Pediatrics,* 146.

MENDEL, M. I. & GOLDSTEIN, R. 1969. Stability of the early components of the averaged electroencephalic response. *Journal of Speech, Language, and Hearing Research,* 12**,** 351-361.

MESGARANI, N., DAVID, S. V., FRITZ, J. B. & SHAMMA, S. A. 2009. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology,* 102**,** 3329-39.

MESIK, J. & WOJTCZAK, M. 2022. The effects of data quantity on performance of temporal response function analyses of natural speech processing. *Frontiers in Neuroscience,* 16**,** 963629.

MIRKOVIC, B., DEBENER, S., SCHMIDT, J., JAEGER, M. & NEHER, T. 2019. Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research,* 377**,** 260-270.

MONTOYA-MARTÍNEZ, J., VANTHORNHOUT, J., BERTRAND, A. & FRANCART, T. 2021. Effect of number and placement of EEG electrodes on measurement of neural tracking of speech. *PLoS One,* 16**,** e0246769.

MOORE, B. C. 2008. The choice of compression speed in hearing AIDS: theoretical and practical considerations and the role of individual differences. *Trends in Amplification,* 12**,** 103-12.

MOORE, B. C. J. & GLASBERG, B. R. 2007. Modeling binaural loudness. *The Journal of the Acoustical Society of America,* 121**,** 1604-1612.

MUNRO, K. J., PURDY, S. C., AHMED, S., BEGUM, R. & DILLON, H. 2011. Obligatory cortical auditory evoked potential waveform detection and differentiation using a commercially available clinical system: HEARLab™. *Ear Hear,* 32**,** 782-6.

MUNRO, K. J., PURDY, S. C., UUS, K., VISRAM, A., WARD, R., BRUCE, I. A., MARSDEN, A., STONE, M. A. & VAN DUN, B. 2020. Recording Obligatory Cortical Auditory Evoked Potentials in Infants: Quantitative Information on Feasibility and Parent Acceptability. *Ear and Hearing,* 41**,** 630-639.

NÄÄTÄNEN, R., KUJALA, T. & WINKLER, I. 2011. Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology,* 48**,** 4-22.

NATIONAL HEALTH SERVICE (NHS). 2021. *Newborn hearing screening* [Online]. The National Health Service. Available: https://www.nhs.uk/conditions/baby/newborn-screening/hearing-test/ [Accessed 22 Aug 2022].

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE (NICE) 2018. 16, Hearing aid microphones and noise reduction algorithms. *Hearing loss in adults: assessment and management*. London: National Guideline Centre (UK).

NIQUETTE, P., ARCAROLI, J., REVIT, L., PARKINSON, A., STALLER, S., SKINNER, M. & KILLION, M. 2003. Development of the BKB-SIN Test. Annual meeting of the American Auditory Society, 12-15 March 2003 Scottsdale, AZ.

NOVIS, K. & BELL, S. 2019. Objective Comparison of the Quality and Reliability of Auditory Brainstem Response Features Elicited by Click and Speech Sounds. *Ear and Hearing,* 40**,** 447-457.

O'SULLIVAN, J. A., POWER, A. J., MESGARANI, N., RAJARAM, S., FOXE, J. J., SHINN-CUNNINGHAM, B. G., SLANEY, M., SHAMMA, S. A. & LALOR, E. C. 2015. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex,* 25**,** 1697-706.

OKAMOTO, H., TEISMANN, H., KAKIGI, R. & PANTEV, C. 2012. Auditory evoked fields elicited by spectral, temporal, and spectral-temporal changes in human cerebral cortex. *Frontiers in Psychology,* 3**,** 1-11.

ORTIZ BARAJAS, M. C., GUEVARA, R. & GERVAIN, J. 2021. The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience,* 48**,** 100915.

OTICON. 2023. *Speech Guard E* [Online]. Available: https://www.oticon.com/professionals/brainhearing-technology/brainhearing-technology/speech-guard-e [Accessed 26 Jun 2023].

PASLEY, B. N., DAVID, S. V., MESGARANI, N., FLINKER, A., SHAMMA, S. A., CRONE, N. E., KNIGHT, R. T. & CHANG, E. F. 2012. Reconstructing speech from human auditory cortex. *PLOS Biology,* 10**,** e1001251.

PÉREZ-GONZÁLEZ, D. & MALMIERCA, M. S. 2014. Adaptation in the auditory system: an overview. *Frontiers in Integrative Neuroscience,* 8**,** 1-19.

PETERSEN, E. B., WÖSTMANN, M., OBLESER, J. & LUNNER, T. 2017. Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology,* 117**,** 18-27.

PICTON, T. W., GOODMAN, W. S. & BRYCE, D. P. 1970. Amplitude of Evoked Responses to Tones of High Intensity. *Acta Oto-Laryngologica,* 70**,** 77-82.

PICTON, T. W. & HILLYARD, S. A. 1974. Human auditory evoked potentials. II. Effects of attention. *Electroencephalography and Clinical Neurophysiology,* 36**,** 191-9.

PICTON, T. W., HILLYARD, S. A., KRAUSZ, H. I. & GALAMBOS, R. 1974. Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology,* 36**,** 179-190.

POWER, A. J., FOXE, J. J., FORDE, E. J., REILLY, R. B. & LALOR, E. C. 2012. At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience,* 35**,** 1497-1503.

PRAKASH, H., ABRAHAM, A., RAJASHEKAR, B. & YERRAGUNTLA, K. 2016. The effect of intensity on the speech evoked auditory late latency response in normal hearing individuals. *Journal of International Advanced Otology,* 12**,** 67-71.

PRINSLOO, K. D. & LALOR, E. C. 2022. General Auditory and Speech-Specific Contributions to Cortical Envelope Tracking Revealed Using Auditory Chimeras. *Journal of Neurophysiology,* 42**,** 7782-7798.

PUNCH, S., VAN DUN, B., KING, A., CARTER, L. & PEARCE, W. 2016. Clinical Experience of Using Cortical Auditory Evoked Potentials in the Treatment of Infant Hearing Loss in Australia. *Seminars in Hearing,* 37**,** 36-52.

PURDY, S. & KELLY, A. 2001. Cortical auditory evoked potential testing in infants and young children. *The New Zealand Audiological Society Bulletin,* 11**,** 16-24.

PURVES, D. A., G. J.; FITZPATRICK, D. 2001. The External Ear. *Neuroscience*. Sunderland (MA): Sinauer Associates.

QUIROGA, R. Q. & GARCIA, H. 2003. Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology,* 114**,** 376-390.

RIGGINS, T., SCOTT, L. & NELSON, C. 2007. Methods for acquiring and analysing infant event-related potentials. *Infant EEG and Event-Related Potentials*.

ROEBUCK, H. & BARRY, J. G. 2018. Parental perception of listening difficulties: an interaction between weaknesses in language processing and ability to sustain attention. *Scientific Reports,* 8**,** 6985.

ROSEN, S., CARLYON, R. P., DARWIN, C. J. & RUSSELL, I. J. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 336**,** 367-373.

ROTHMAN, H. H. 1970. Effects of High Frequencies and Intersubject Variability on the Auditory-Evoked Cortical Response. *The Journal of the Acoustical Society of America,* 47**,** 569-573.

RUHM, H., WALKER JR., E. & FLANIGIN, H. 1967. Acoustically-evoked potentials in man: Mediation of early components. *The Laryngoscope,* 77**,** 806-822.

SCOLLIE, S., GLISTA, D., TENHAAF, J., DUNN, A., MALANDRINO, A., KEENE, K. & FOLKEARD, P. 2012. Stimuli and normative data for detection of Ling-6 sounds in hearing level. *American Journal of Audiology,* 21**,** 232–241.

SHANNON, R. V., ZENG, F. G., KAMATH, V., WYGONSKI, J. & EKELID, M. 1995. Speech recognition with primarily temporal cues. *Science,* 270**,** 303-304.

SKOE, E. & KRAUS, N. 2010. Auditory brain stem response to complex sounds: a tutorial. *Ear and Hearing,* 31**,** 302-24.

SOUZA, P. 2016. Speech Perception and Hearing Aids. *In:* POPELKA, G. R., MOORE, B. C. J., FAY, R. R. & POPPER, A. N. (eds.) *Hearing Aids*. Cham: Springer International Publishing.

SOUZA, P., HOOVER, E. & GALLUN, F. 2012. Application of the envelope difference index to spectrally sparse speech. *Journal of Speech, Language, and Hearing Research,* 55**,** 824-37.

SOUZA, P. E. 2002. Effects of Compression on Speech Acoustics, Intelligibility, and Sound Quality. *Trends in Amplification,* 6**,** 131-165.

STONE, M. A. & MOORE, B. C. 2007. Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America,* 121**,** 1654-64.

SUMMERFIELD, C., TRITTSCHUH, E. H., MONTI, J. M., MESULAM, M. M. & EGNER, T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience,* 11**,** 1004-6.

TAYLOR, B. 2007. *Predicting Real World Hearing Aid Benefit with Speech Audiometry: An Evidence-Based Review* [Online]. Audiology online. Available: https://www.audiologyonline.com/articles/predicting-real-world-hearing-aid-946 [Accessed 20 Aug 2020].

UNDE, S. A. & SHRIRAM, R. Coherence Analysis of EEG Signal Using Power Spectral Density. 2014 Fourth International Conference on Communication Systems and Network Technologies, 7-9 April 2014 2014. 871-874.

VAN DUN, B., CARTER, L. & DILLON, H. 2012. Sensitivity of Cortical Auditory Evoked Potential Detection for Hearing-Impaired Infants in Response to Short Speech Sounds. *Audiology Research,* 2**,** e13.

VAN HIRTUM, T., SOMERS, B., DIEUDONNÉ, B., VERSCHUEREN, E., WOUTERS, J. & FRANCART, T. 2023. Neural envelope tracking predicts speech intelligibility and hearing aid benefit in children with hearing loss. *Hearing Research,* 439**,** 108893.

VAN TASELL, D. J. & YANZ, J. L. 1987. Speech Recognition Threshold in Noise. *Journal of Speech, Language, and Hearing Research,* 30**,** 377-386.

VANHEUSDEN, F. J., CHESNAYE, M. A., SIMPSON, D. M. & BELL, S. L. 2019. Envelope frequency following responses are stronger for high-pass than low-pass filtered vowels. *International Journal of Audiology,* 58**,** 355-362.

VANHEUSDEN, F. J., KEGLER, M., IRELAND, K., GEORGA, C., SIMPSON, D. M., REICHENBACH, T. & BELL, S. L. 2020. Hearing Aids Do Not Alter Cortical Entrainment to Speech at Audible Levels in Mild-to-Moderately Hearing-Impaired Subjects. *Frontiers in Human Neuroscience,* 14**,** 109.

VANTHORNHOUT, J., DECRUY, L. & FRANCART, T. 2019. Effect of Task and Attention on Neural Tracking of Speech. *Frontiers in Neuroscience,* 13**,** 977.

VANTHORNHOUT, J., DECRUY, L., WOUTERS, J., SIMON, J. Z. & FRANCART, T. 2018. Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *Journal of the Association for Research in Otolaryngology,* 19**,** 181-191.

VARNET, L., ORTIZ-BARAJAS, M. C., ERRA, R. G., GERVAIN, J. & LORENZI, C. 2017. A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America,* 142**,** 1976-1989.

VERSCHUEREN, E., GILLIS, M., DECRUY, L., VANTHORNHOUT, J. & FRANCART, T. 2022. Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech Rate Manipulation Paradigm. *The Journal of Neuroscience,* 42**,** 7442.

VERSCHUEREN, E., VANTHORNHOUT, J. & FRANCART, T. 2021. The effect of stimulus intensity on neural envelope tracking. *Hearing Research,* 403**,** 108175.

VERSCHUURE, J., MAAS, A. J. J., STIKVOORT, E., DE JONG, R. M., GOEDEGEBURE, A. & DRESCHLER, W. A. 1996. Compression and its Effect on the Speech Signal. *Ear and Hearing,* 17**,** 162-175.

VISRAM, A. S., STONE, M. A., PURDY, S. C., BELL, S. L., BROOKS, J., BRUCE, I. A., CHESNAYE, M. A., DILLON, H., HARTE, J. M., HUDSON, C. L., LAUGESEN, S., MORGAN, R. E., O'DRISCOLL, M., ROBERTS, S. A., ROUGHLEY, A. J., SIMPSON, D. & MUNRO, K. J. 2023. Aided Cortical Auditory Evoked Potentials in Infants With Frequency-Specific Synthetic Speech Stimuli: Sensitivity, Repeatability, and Feasibility. *Ear and Hearing,* 44**,** 1157-1172.

VONCK, B. M. D., LAMMERS, M. J. W., VAN DER WAALS, M., VAN ZANTEN, G. A. & VERSNEL, H. 2019. Cortical Auditory Evoked Potentials in Response to Frequency Changes with Varied Magnitude, Rate, and Direction. *Journal of the Association for Research in Otolaryngology,* 20**,** 489-498.

WANG, L., WU, E. X. & CHEN, F. 2020. Robust EEG-based decoding of auditory attention with high-rms-level speech segments in noisy conditions. *Frontiers in Human Neuroscience,* 14**,** 557534.

WORLD HEALTH ORGANIZATION. 2021. *Deafness and hearing loss* [Online]. World Health Organization. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [Accessed 20 Aug 2022].

ZOEFEL, B. & VANRULLEN, R. 2016. EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage,* 124**,** 16-23.

ZOU, J., FENG, J., XU, T., JIN, P., LUO, C., ZHANG, J., PAN, X., CHEN, F., ZHENG, J. & DING, N. 2019. Auditory and language contributions to neural encoding of speech features in noisy environments. *NeuroImage,* 192**,** 66-75.