University of Southampton

Faculty of Engineering and Physical Sciences
School of Engineering

Speech-in-noise performance in hearing-impaired listeners assessed using evoked responses and enhanced using tactile stimulation

by

Samuel William Perry
BA, MSc
ORCiD: 0000-0001-5925-6332

A thesis for the degree of
Doctor of Philosophy

November 2022

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Engineering

Doctor of Philosophy

Speech-in-noise performance in hearing-impaired listeners assessed using evoked responses and enhanced using tactile stimulation

by Samuel William Perry

Hearing aid and cochlear implant users struggle to understand speech in noisy places, such as classrooms and busy workplaces, with their performance typically being significantly worse than for normal-hearing listeners. This thesis details development of two new methods for improvement of speech-in-noise performance outcomes. The first addresses shortcomings in current techniques for assessing speech-in-noise performance and the second proposes a new intervention to improve performance.

Chapters 3 and 4 present modifications to a new electrophysiological assessment method, using the temporal response function (TRF), for prediction of speech-in-noise performance. The TRF offers information not provided by behavioural speech-in-noise measures (the gold standard for speech-in-noise research and clinical assessment), which may be used for automated intervention fitting and further analysis of the mechanisms of speech-in-noise performance. Alterations to methodology for applying the TRF are proposed, which may provide the groundwork for further development of the TRF as a method for assessing speech-in-noise performance.

Chapters 5 and 6 investigate the efficacy of a new intervention to improve speech-in-noise performance in cochlear implant users by providing missing sound-information through tactile stimulation on the wrists. This section focuses on developing and testing initial prototype devices that could rapidly be adapted for real-world use. These prototypes represent the first step towards the realisation of a wearable device, with accompanying results demonstrating the potential for their use in improving speech-in-noise performance.

This thesis highlights two techniques that could be further developed for assessing and enhancing speech-in-noise performance, and outlines future steps to be taken for the realisation and combination of these techniques for improved treatment of the hearing impaired.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is 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: M. D. Fletcher et al. (Dec. 2020c). "Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic". In: Scientific Reports 10.1. issn: 2045-2322. doi: 10.1038/s41598-020-67140-0. url: http://www.nature.com/articles/s41598-020-67140-0 (visited on 07/29/2020)

Signed:.. Date:..................

Acknowledgements

I would like to thank my supervisors Prof. Steve Bell and Prof. David Simpson for their enthusiastic support and guidance throughout the project. I would also like to thank my supervisor Dr. Mark Fletcher for stepping in during difficult times. His insights and alternative solutions were invaluable. I also wish to thank Dr Matt Blyth for his clinical expertise, moral support and for generally being a great cheerleader. Thanks also to Dr. Freddy Vanheusden for his guidance with all things EEG. I would like to thank Robyn Hunt for her advice and support throughout the process. I wish to thank P.T Chaplin for taking the time to tutor me in analog electronics. I also thank Chris Rayment for countless hours of technical discussion and IT support.

I acknowledge the funders of this project: EPSRC and The William Demant Foundation. I would like to acknowledge the contributions of University of Southampton Auditory Implant Service and its staff for additional funding and support during the project. I would also like to thank the participants who volunteered their time for my studies.

I would like to thank my family: Debs, Dave, Eloise, Tom and Wally; for their love, support and patience. To them, I also apologise for the considerable stress my academic journey has caused them and promise to finally get a real job. I would also like to thank my extended northern family - in particular Sam Prieditis, Vicki Prieditis, Will Shannon and Sarah Shannon, for offering refuge and for generally putting up with me. Without them I would not have made it to where I am. Finally, I would like to thank my partner, Shuchi Purohit, for her love and support, for keeping me well fed from near or afar, and for her untiring patience through the many stresses over the years.

Nomenclature

f_0 Fundamental frequency, page 21

ABR Auditory brainstem response, page 29

ALR Auditory late response, page 29

AMLR Auditory mid-latency response, page 29

ASIC Application-specific integrated circuit, page 42

ASSR Auditory steady state response, page 29

B&K Brüel & Kjær, page 116

BKB Bamford-Kowal-Bench, page 19

cABR Complex auditory brainstem response, page 35

DSS Denoising source separation, page 38

ECochG Electrocochleography, page 8

EEG Electroencephalography, page 8

FFR Frequency following response, page 36

HINT Hearing in noise test, page 19

LTASS Long term average speech spectrum, page 24

MEG Magnetoencephalography, page 8

MMN Mismatch negativity response, page 35

mTRF Multivariate temporal response function, page 39

PTA Pure Tone Audiometry, page 1

SNR Signal to noise ratio, page 18

SRT Speech reception threshold, page 18

SSQ Speech, spatial and quality of hearing scale, page 36

TRF Temporal response function, page 39

TRL Technology Readiness Level, page 141

USAIS The University of Southampton's Auditory Implant Service, page 25

WIN Word in noise, page 19

Chapter 1

Introduction

Outside of the laboratory, speech is rarely encountered in isolation. It is typically experienced in suboptimal conditions, competing with a variety of background noises. As one of the most important acoustic signals encountered in everyday life, a person's ability to understand speech in background noise can have a significant impact on their ability to communicate (Carhart and Tillman, 1970; Dubno et al., 1984; Mattys et al., 2012), affecting academic ability, workplace performance and social interactions (Shield and Dockrell, 2008; Sullivan and Carrano, 2015; Dobie and Van Hemel, 2004). Poor speech-in-noise performance is one of the most common complaints of individuals with hearing loss (Kochkin, 2000; Plomp, 1986; Dubno et al., 1984) and is typically caused by pathologies or lesions in the auditory system (as described in Section 2.1). High quality speech-in-noise performance assessment and interventions are therefore key to achieving improved benefits for patients with hearing loss.

Clinical assessment of speech-in-noise typically relies on behavioural testing to measure a patient's performance. Patients are assessed by playing speech stimuli in background noise and measuring their ability to repeat what was heard. This provides a quick and easy to administer method to assess the benefits of a hearing assistive device. These tests also offer additional insight to that of standard tests such as pure tone audiometry (PTA), by providing a more direct measure of the communication difficulties experienced by the patient (Taylor, 2003). However, there are also several limitations to these tests. First, they provide little information about the underlying cause of speech-in-noise performance issues (such as whether the issue is as a result of a conductive or senorineural hearing loss — see Section 2.1 for details). This limits their use in diagnosis of hearing issues, as additional tests are required to determine the cause. Speech-in-noise test results also offer minimal indication of how to adjust an intervention to better suit the patient, beyond the patient's own subjective responses. Again, additional tests (such as PTA) are therefore required for the fitting procedure, limiting speech-in-noise tests to be used primarily for verification of a fitting. Finally, speech-in-noise tests require active participation from the patient, making them unsuitable for certain groups of patient, such as children and those unable to respond (Anderson and Kraus, 2010).

Evoked responses are an alternative objective method that measures synchronous electrical neural activity in response to sensory stimulus. They have not currently been developed for clinical assessment of speech-in-noise performance, but have been demonstrated as a potential alternative measure (Vanthornhout et al., 2018; Anderson and Kraus, 2010).

Evoked responses have a number of benefits over behavioural measures that make them a viable candidate for an alternative measure of speech-in-noise performance. First, they don't require direct user interaction, so may be developed as an appropriate measure for those who are unable to respond behaviourally. This may be of particular benefit for assessing infants, in order to appropriately fit a hearing intervention earlier than is possible with behavioural measures. Evoked responses may also provide additional insight on the neural mechanisms that affect speech-in-noise performance. Additional information may lead to more informed optimisation of current hearing assistive devices. For example, a better understanding of the types of acoustic features that are being poorly transmitted along the auditory pathways, may allow clinicians to tune the patient's assistive device accordingly.

Evoked responses also have application in development of new alternative interventions for addressing speech-in-noise issues issues (Ding and Simon, 2014; Vanthornhout et al., 2018). By embedding electrodes in hearing assistive technology, it may be possible to measure responses in real-time. This would allow for on-the-fly tuning of hearing interventions based on the neural activity of the patient.

There are a number of limitations that should be addressed in the realisation of evoked responses as a viable clinical speech-in-noise measure. In particular the testing time needed to record evoked responses is considerably longer than that for behavioural measures. This will need to be reduced for a clinical implementation of any proposed evoked response methods, given the limited time available for hearing intervention fitting in clinic. Evoked response based methods are also more cumbersome, using multiple electrodes in addition to the headphones/loudspeaker and button required for behavioural measures. This will also need to be reduced for clinical implementations, as use of many electrodes may cause issues, particularly with uncooperative patients such as infants. Finally, it should be noted that otoacoustic emissions are an alternative measure of hearing performance that offer objectively assess outer and middle ear function. This technique is used widely in clinics, in particular for assessment of newborns (Akinpelu et al., 2014). The specific mechanism (cochlear amplification) that this technique measures limit's their applicability in providing a complete speech-in-noise assessment method. For these reasons, this thesis will focus on the application of evoked responses, given their ability to assess the auditory system from brainstem to cortex (as detailed in the following Section 2.2.2).

In addition to the effective measurement of speech-in-noise performance, an effective intervention must be provided to address the patient's performance issues. Current clinical interventions, such as hearing aids and cochlear implants, are typically employed to counter these issues. Whilst these interventions are effective in many cases, they do not provide a complete solution, with some patients still experiencing significant issues (Fagan, 2015; Wilson, 2008). This is partially due to the biological limitations of the damaged auditory system and the technological limitations of the interventions. Research into the use of multi-sensory interventions, such as electro-haptic stimulation (a technique that combines the electrical stimulation of a cochlear implant with haptic stimulation provided on a patient's wrists) and haptic aids, is a potential alternative to address some of the limitations of current interventions. These methods aim to provide information not transmitted by hearing based intervention, via alternative senses such as touch. It has been suggested that by augmenting hearing with haptic stimulation, it may be possible to improve a listener's ability to recognise speech in background noise (Fletcher, 2021a; Fletcher et al., 2020b; Fletcher et al., 2018).

This thesis has two primary aims. The first is to explore the efficacy of evoked responses for assessment of speech-in-noise performance. The second is to explore the efficacy of electro-haptic stimulation for treatment of speech-in-noise performance issues. Combined, these areas have the potential to provide the basis for a new, objectively optimised alternative to current hearing assistive devices.

Chapter 2

Background

This chapter outlines current research relevant to the development of a new electrophysiologically -based speech-in-noise assessment and haptics-based intervention. Section 2.1 will first outline the physiology of the auditory system that influences a person's speech in noise performance, as well as the forms of pathologies and lesions that may lead to a reduction in a person's speech-in-noise performance. The relative merits of behavioural and electrophysiological measures of speech-in-noise performance are discussed in detail in Section 2.2. Key considerations will be outlined for the development of both practical clinical tests, as well as tests for in-lab research. Section 2.3 will then outline current clinical interventions, such as hearing aids and cochlear implants which are employed to counter issues with speech-in-noise performance. This section will highlight the benefits of these interventions, and also highlight the practical and technical challenges that limit their performance. Finally, Section 2.4 will introduce the use of haptics as a further form of intervention, which may be used to augment current clinical interventions. This section will outline current understanding of the integration of haptic and auditory stimulus for the improvement of speech-in-noise performance. The technical and practical challenges currently faced when implementing a real-world intervention using this technology will also be outlined.

2.1 Mechanisms of speech-in-noise performance

A person's ability to hear speech in background noise depends on their sensory system's ability to separate relevant acoustic and linguistic cues from noise. This complex process is performed primarily by the auditory system, but can also be supported through integration with other sensory modalities, such as vision and touch. Section 2.1.1 provides an overview of the anatomy and physiology of the auditory system systems and outlines interactions with other modalities when faced with a speech-in-noise task. Section 2.1.2 covers the key perceptual features of speech and types of noise and distortion.

Finally, Section 2.1.3 outlines various hearing pathologies that have a negative impact on speech-in-noise performance.

2.1.1 Anatomy and physiology of audition for speech recognition

The human auditory system consists of a number of components that work together to process sound signals. The auditory system can be divided into a number of components, which are described in the following sections: The conductive system (Section 2.1.1.1), consisting of the outer and middle ear; The sensorineural system (Section 2.1.1.2), consisting of the cochlea and eighth cranial nerve; and the central auditory pathways (Section 2.1.1.3), comprising of a complex network of neurons, nerves and nuclei that lead from the cochlea to the auditory cortex. Additional systems such as the visual and tactile system, and their integration with the auditory system are detailed in Section 2.1.1.5. Finally, computational models of the processes that occur in the aforementioned sensory systems are detailed in Section 2.1.1.6.

2.1.1.1 The outer and middle ear

The outer and middle ear's primary function is to convert the acoustic pressure changes in the air to fluid vibrations in the cochlea. Sound signals are first filtered by the physical shape of the listener's head, the pinna and resonances in the ear canal. These are then converted to mechanical vibrations at the eardrum (tympanic membrane), which travel via the ossicles (the Malleus, Incus and Stapes) to the cochlea (Moore, 2016; Plack, 2014, p.24, p.53-55). The anatomy of the outer ear is illustrated in Figure 2.1.

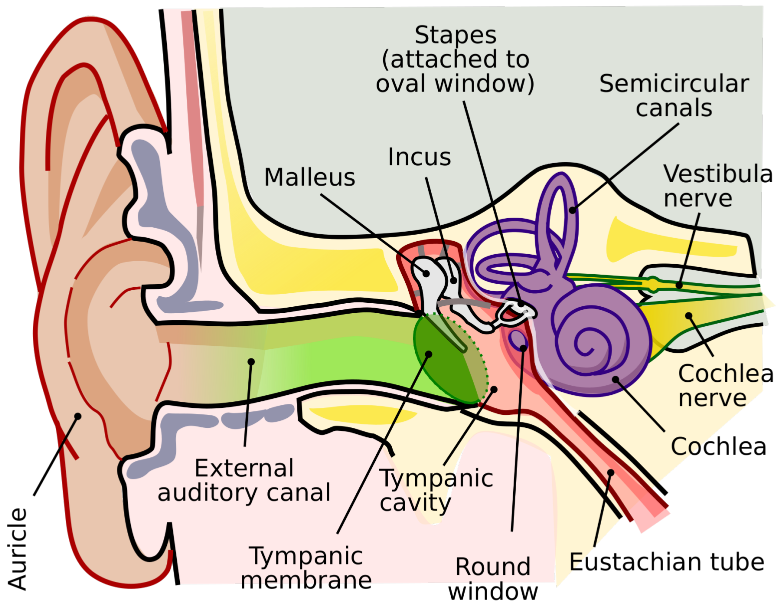

Figure 2.1: Schematic diagram of the peripheral auditory system[1]

These mechanical processes filter the sound, altering its temporal and spectral characteristics prior to transduction in the inner ear. Damage to the outer or middle ear can distort speech and noise signals, which may impact a person's speech-in-noise performance. This is known as a "conductive hearing loss" — a brief overview on the types of hearing loss is provided in Section 2.1.3.

2.1.1.2 The inner ear

The inner ear is concerned with the conversion of acoustic vibrations, transmitted by the outer ear, into neural signals to be further processed by the central auditory pathways. This process of transduction occurs in the cochlea — a coiled, fluid filled tube, roughly 3.5cm in length. The primary components of the cochlea in relation to sound transduction are the basilar membrane, a thin membrane that spans the length of the cochlea; the organ of corti, a layer of tissue which rests on top of the basilar membrane; and outer and inner hair cells, which line the organ of corti. Cavities on both side of the basilar membrane vary in pressure, caused by the vibration of the oval window (a membrane covered opening in the cochlea, connected to the stapes). Vibration of the basilar membrane, as a result of changes in fluid pressure and the movement of outer hair cells, causes movement of inner hair cells. In turn the movement of these hair cells triggers the firing of neurons, thus converting the acoustic vibration to neuronal activity. This activity is then subject to further processing in the central auditory pathways. The detailed workings of the cochlea can be found in texts such as those provided by Gelfand (2016) and Moore (2016) and Plack (2014).

In addition to the transduction of vibrations to neural activity, the cochlea is of particular importance to speech perception, as it is the point at which sound is separated based on its spectral characteristics. By mapping specific frequency sub-bands of sounds to neurons along the length of the basiliar membrane, the cochlea creates a tonotopic neural representation of a sound. This process effectively behaves analogously to a set of band-pass filters, which act to divide the acoustic vibrations into bands that increase from low to high frequency. These filters are commonly referred to as "auditory filters" (Moore, 2016; Plack, 2014, p.68, p.57-59). This tonotopic mapping is an important stage in auditory processing and is thought to be a key component in fundamental processes of speech recognition, such as pitch perception, formant tracking and loudness perception (Plack, 2014, p.100, 117).

[1]This creative commons licensed figure was produced by Lars Chittka and Axel Brockmann (`https://commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg`), "Anatomy of the Human Ear", `https://creativecommons.org/licenses/by/2.5/legalcode` (accessed 05/12/2021).

2.1.1.3 The central auditory pathways

Beyond the cochlea, acoustic information is carried as electrical activity via structures of connected neurons through the central auditory pathway. This section aims to provide an overview of the components and structure of the central auditory pathways. The function of individual cells and their combined interactions in higher level structures is described in relation to speech perception. However, it should be noted that the auditory pathways are not fully understood due to their significant complexity. Therefore, this section provides a brief overview of the neurophysiological aspects and systems, thought to be of particular importance in the processing of speech in noise. For further information on the details of neurophysiology, a comprehensive overview of current knowledge with regards to the central nervous system as a whole (of which the central auditory pathway is a significant component), is provided by Bear et al. (2016). Schnupp et al. (2011) also provides significant detail on the neural processing of sound and the perceptual correlates of these processes.

The central auditory pathways consist of two types of cell: neurons and glia. Auditory signals are processed via structures of connected neurons as they ascend the auditory pathway, with surrounding glia cells thought to primarily support and maintain neurons (Bear et al., 2016). Neurons transmit information as electrical spikes known as action potentials. These action potentials travel from neuron to neuron via axons, which synapse with the soma, axons and dendrites of other neurons to form a network. Information such as low level audio feature representations are encoded in the frequencies and patterns of these potentials. Examples include amplitude modulations and spectral features of a stimulus, which are encoded by phase locking the spikes produced by neurons to the frequency of the feature (Plack, 2014; Schnupp et al., 2011). In addition, as action potentials cause synaptic excitation as signals are passed between neurons, secondary extracellular currents are generated which are known as field potentials. These potentials propagate to the scalp as extracellular activity in the neurons and glia. This activity can be measured using techniques such as Electroencephalography (EEG), Magnetoencephalography (MEG) and Electrocochleography (ECochG). Use of these techniques for analysis of neural activity in response to acoustic stimuli and for the potential analysis of speech in noise performance is discussed in Section 2.2.2. As acoustic signals ascend the auditory system they are passed between various structures of neurons called nuclei. These nuclei are connected both in ascending connections and descending efferent pathways, sending auditory signals upwards to subsequent nuclei and feedback signals downward, thought to affect previous auditory processing mechanisms (Burguetti and Carvallo, 2008). It is thought that the auditory pathway is formed in a hierarchical structure that performs low to high level processing similar to that of the visual system (Moore, 2016; Okada et al., 2010). The pathways can be separated into three main divisions: the brainstem, thalamus and cerebrum. Each division is involved to various degrees in the low, mid and high level processing that is performed on speech signals (Gelfand, 2018). A schematic diagram of the primary neural connections in the system is provided in Figure 2.2.

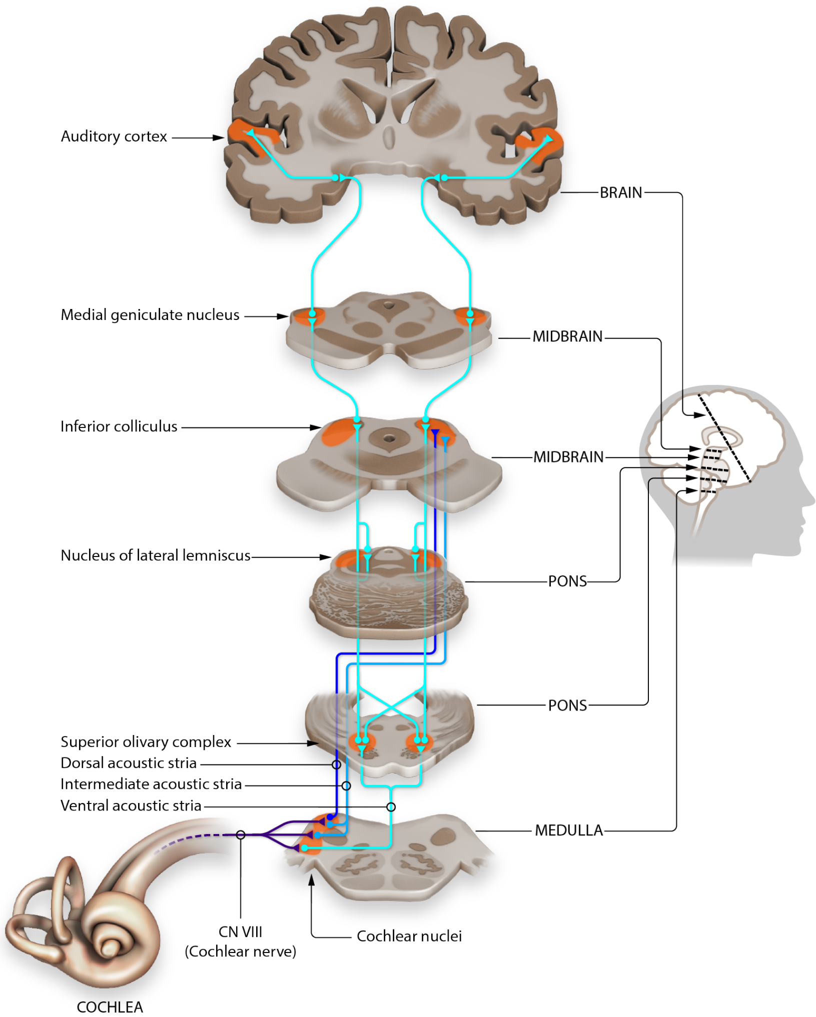

Figure 2.2: Schematic diagram of the central auditory pathways[2]

The central auditory pathways consist of a vast network of interconnecting neurons that carry electrical signals, generated at the cochlea, to a variety of processing stations.

Initially, information travels through the brainstem to the Cochlear Nucleus via the auditory nerve, at which point signals are distributed via branches of neurons to subsequent auditory processing stations (nuclei). Nuclei in the brainstem that are believed to be of particular importance in speech processing include: the superior olivary complex, the point at which a significant amount of information from both ears converge and

[2]This creative commons licensed figure was produced by Jonathan E. Peelle (https://commons.wikimedia.org/wiki/File:Auditory_Pathway.png), "Auditory Pathway", https://creativecommons.org/licenses/by/4.0/legalcode (accessed 05/12/2021).

therefore thought to play a key role in sound spacialisation; the nucleus of lateral lemniscus, which has been hypothesized as a source of acoustic reverberation processing, as well as in the shaping of interaural time differences used for sound localisation (Felmy, 2019; Kidd and Kelly, 1996); and the inferior colliculus, which is thought to perform processing related to the periodicity of the stimulus (Schnupp et al., 2011, p.125-129). At this stage, tonotopic mapping of neurons is thought to be maintained, and it has been shown that neurons respond directly to simple stimuli such as tones (Hickok and Poeppel, 2007; Binder, 2000; Wessinger et al., 2001). This suggests that the brainstem is largely concerned with processing of lower level acoustic features (features such as the fundamental frequency or amplitude modulations of a stimulus, which are related to its acoustic characteristics). However, the brainstem may also play a role in higher-level linguistic processing, due to the extensive efferent pathways that link the brainstem to higher centres of the brain (Budinger et al., 2000; Doucet et al., 2002; Coomes and Schofield, 2004). For example, recent research using electroencephalography (EEG) has suggested that the response of neurons in the brainstem may also be influenced by higher level processing such as selective attention (Forte et al., 2017; Reichenbach et al., 2016). Evidence has also been presented that suggests representations of higher level features derived from pitch may be present in subcortical mechanisms (Krishnan and Gandour, 2009).

As signals ascend the auditory pathway, they reach the medial geniculate body in the thalamus. As with the brainstem, a tonotopic arrangement of neurons is thought to be maintained at this stage. However, it is thought that representations of speech signals no longer accurately entrain the rhythm of speech (Schnupp et al., 2011, p.155-156). It has also been shown that speech processing in the medial geniculate body is task dependant and influenced by attention (von Kriegstein et al., 2008). This influence of higher level factors and divergence of signal representations from the fine grained rhythmic structures of speech illustrates the transition from lower level acoustic signal processing towards higher level linguistically-relevant representations as signals ascend through the auditory pathway. This continues as signals travel on to the auditory cortex.

The auditory cortex is an auditory processing station located in the cerebrum. Situated in the temporal lobe, the auditory cortex, as well as auditory processing centres known as the planum polare and planum temporale, interact with higher order cognitive structures (Schnupp et al., 2011). Aspects of particular interest with regards to speech recognition are the superior temporal sulcus, the superior temporal gyrus, and Wernicke's area. Situated primarily in the left hemisphere, clinical observation has linked this area to speech production and comprehension. This has been shown through observation of temporary disruption to these components, which can have a negative impact on speech processing tasks (Bear et al., 2016). It has also been shown through experiments using positron emission tomography (PET) scans, that sections of the superior temporal gyrus respond with preference to intelligible speech over unintelligible speech (Scott, 2000). In the primary sections of cortical processing, it is thought that signals are still organised by their acoustic features (Schnupp et al., 2011, p.165). It is believed that as the signals disperse through the cortex, passing through areas such as the superior temporal gyrus towards the prefrontal cortex in the frontal lobe, processing is concerned with the mapping of acoustic features to categorical lexical and semantic representations (Stevens, 2002; Wilson et al., 2018). At the cortical stage of processing, cognitive functions such as working memory and attentional capacity begin to significantly influence speech processing. For example, working memory has been shown to affect speech recognition performance, particularly in its role as temporary storage for unidentified words until they can be successfully parsed using subsequent contextual cues. This is of importance when a word cannot be immediately identified for reasons such as lack of clear articulation by a speaker (Wingfield, 1996). The effects of higher-level linguistic processing mechanisms on speech recognition performance are discussed further in Section 2.1.2. However, higher-level linguistic processes are not fully understood, in part due to the limitations of neuroimaging approaches. Reasons for this are discussed briefly in Section 2.2.2. This presents a significant challenge when considering speech-in-noise performance: Without a clear understanding of higher-level linguistic recognition and semantic processing in the brain, it is difficult to understand the effect of noise on processes that are crucial to speech perception and recognition. Current understanding of the effects that noise has on speech perception and of the methods used to measure performance in this area are discussed in the following sections.

2.1.1.4 Binaural hearing

The arrival time and level differences (interaural time differences and interaural level differences) of a signal between two ears are known to play a key role in sound localisation. This has been shown to affect speech recognition in noise, in particular when the target speaker and distracting noise are not co-located. This ability of a listener to focus their attention on a target speaker in the presence of competing, spatially separate noise is often referred to as the “cocktail party” effect, particularly in relation to a person’s ability to understand speech in competing, spatially separated babble noise (Cherry, 1953). It has been reported that this spatial release from masking can improve speech reception thresholds significantly when the noise is spatially separated (Bronkhorst and Plomp, 1988). Work such as that produced by Feuerstein (1992) also demonstrates a reduction in word recognition performance for spatially separated speech and noise, as a result of a simulated single-sided unilateral conductive hearing loss in normal hearing participants. This is further supported by Hsieh et al. (2009), through an observed reduction in speech-in-noise performance in unilateral conductive hearing loss participants. However, due to the greater impact of hearing deficits in other auditory systems (as discussed in the following sections), this research will focus specifically on speech-in-noise performance in a monaural listening environment. It should therefore be noted that this will negate some of the most prominent effects of deficits in the outer ear, such as changes in filtering as a result of a damaged pinna and ear canal, which may prove to be an interesting area for further research.

2.1.1.5 Multi-sensory integration

In addition to direct auditory stimulation, speech and language recognition can be affected by input from other senses. This ‘multi-sensory integration’ has been shown to provide considerable improvements to speech recognition performance, primarily through audio-visual integration. For example, lip-reading is the integration of auditory and visual cues, and has been shown to support speech understanding in particular when auditory cues are degraded (Grant and Braida, 1991). This integration is thought to occur primarily in the cortical regions, such as the primary auditory cortices, the angular gyrus and the inferoposterior temporal lobe (Bernstein et al., 2008; Calvert et al., 1997; Campbell et al., 2001).

An area that has also shown promise is audio-tactile integration. This area has a growing literature of research suggesting potential benefits, relating to speech recognition, such as sound localisation (Fletcher et al., 2020a), speech-in-noise performance (Fletcher et al., 2020b; Fletcher et al., 2019; Fletcher et al., 2018; Huang et al., 2017), melody recognition (Huang et al., 2019; Luo and Hayes, 2019) and basic auditory feature perception such as pitch (Fletcher et al., 2020c). Details on the anatomy, limits of the tactile system and its integration with the auditory system are discussed further in Section 2.4.1.

The principle of inverse effectiveness is a key principle of multi-sensory integration (Wallace et al., 1996; Hairston et al., 2003; Laurienti et al., 2006). This states that integration of multiple senses is maximised when each sense is degraded alone. Use of multi-sensory integration is therefore appealing for development of interventions for hearing impaired listeners, as audition is degraded both by competing noise and by the hearing impairment, so is likely to be well supported by information for peripheral senses. This is exemplified in previous studies, where cochlear implant users are provided with partial/degraded location or speech information via both a cochlear implant and haptic devices (Fletcher et al., 2020a; Fletcher et al., 2020b).

A further principle is the correlation of temporal properties (Ernst and Bülthoff, 2004; Fujisaki and Nishida, 2005; Burr et al., 2009; Parise and Ernst, 2016). This states that two senses are most likely to be integrated when stimuli are maximally correlated in time. Therefore, this must be considered in the development of any multisensory based intervention to improve speech recognition, and will be discussed in further sections.

Previous literature shows clear potential for multisensory integration to assist hearing impaired listeners who struggle with speech-in-noise performance. This thesis will focus on audio-tactile integration, due to recent technological advancements that may allow for further benefits than previously shown. A review of modern haptic technology and previous haptic devices that have been developed to improve speech recognition and speech-in-noise performance are detailed in Section 2.4.2.

2.1.1.6 Computational models of the auditory system

A range of computational models have been developed to quantify the relationship between the inputs and outputs of the auditory system's components, detailed previously in this chapter. These models are built on observations, which for the auditory system are typically either physical, neuronal or behavioural measures. In the context of speech-in-noise performance, models are typically based on behavioural data, as these are a direct measure of human perception (as opposed to physical and neuronal measures, which provide only indirect correlates of perception). Lower-level models may aim to produce the outputs of specific parts of the system (such as the response of the middle ear, or the basilar membrane). Higher level models are typically constructed from collections of these lower-level models, and aim to predict the response of the system as a whole — quantifying the perceived loudness of a stimulus (BS.1770-2, 2015), or the speech intelligibility of a stimulus (Kates and Arehart, 2021), for example. A comprehensive list of model types is not feasible within the scope of this thesis, however an extensive collection of model implementations can be found in the Auditory Modelling Toolbox (Majdak et al., 2021). Auditory models have application in hearing-impairment evaluation, by allowing for predictions to be made about how a deficit in components of the auditory system affects its overall function (Majdak et al., 2021). They also have technical application, as they can be used to develop and optimise audio processing strategies in hearing-assistive devices, as discussed in Section 2.3, and also in the development of new diagnostic measures, as demonstrated in Chapters 3 and 4.

2.1.2 Components of speech-in-noise performance

The human auditory system processes various acoustic and lexical cues to understand speech. As noise increases these cues can become distorted or masked entirely, causing a reduction in speech recognition. The type of degradation also dictates the influence of various cognitive systems on recognition performance. Section 2.1.2.1 outlines the various lower- and higher-level cues that are used for segmentation of speech. Section 2.1.2.2 details the types of noise and sound degradations that impact speech recognition performance, and the auditory mechanisms that address these.

2.1.2.1 Acoustic and phonemic features of speech

As described in previous sections, as speech signals ascend through the cortex, representations of acoustic features are transformed into categorical phonemic representations. It is thought that these sequences of sequential phonemic units are used to store words in memory, and that there is a hierarchical structure to this storage (Clements, 1985; McCarthy, 1988; Halle and Stevens, 1971). This mapping is a highly non-linear process, as a change in the phonemic content of a word constitutes a new word, but a change in the raw acoustic cues does not necessarily constitute a new word (for example, words can be spoken with different pitches, intensities or accents without changing the word perceived by the listener). A model proposed by Stevens (2002) outlines a mapping of acoustic cues to split acoustic signals into phonological units. This model presents three broad classes of phonological segment: vowels, glides, and consonants. These segments are delineated by variation in the amplitude envelopes of frequency bands (the bandwidth of which are determined relative to the formants of the speech). The proposed differences between segments are:

- Vowels have greater intensity than consonants
- Vowels have generally higher first formant frequencies than consonants
- Vowels have greater intensity in the low- and mid-frequency spectrum than adjacent consonants
- An acoustic discontinuity occurs at the formation and release of a consonant
- A glide also has a higher frequency at the first formant and a reduction in the low- and mid-frequency spectrum, but there is no acoustic discontinuity

For an in-depth description of the segment types and the anatomical mechanisms for their production, refer to Stevens (2002). This model suggests that the broad spectral shape and intensity of a speech signal are important factors for speech recognition. Many behavioural studies support this conclusion, demonstrating participant's ability to recognise speech when fine-grained spectral cues are removed (Shannon et al., 1995; Shannon et al., 1998; Rosen et al., 1992; Xu and Pfingst, 2008).

In addition to phonemic identification, factors such as fundamental frequency (F_0) recognition should be considered, as this conveys information such as emotional expression (prosody) (Hammerschmidt and Jürgens, 2007), voicing (Holt et al., 2001; Whalen et al., 1993) and is also crucial to tonal languages such as Mandarin Chinese, Thai, Vietnamese and Cantonese. Loss of this information, encoded primarily in the temporal structure of the acoustic signal, has been shown to have a detrimental impact on speech understanding, particularly for tonal languages (Brown and Bacon, 2010; Fu et al., 1998b; Xu and Pfingst, 2008).

2.1.2.2 Noise and distortions

There are 3 main types of noise/degradation (masking) that affect a person's ability to recognise speech: Energetic masking, degradation without energetic masking and informational masking. Energetic masking is formed of environmental masking that additively overwhelms the target speech source due to the intensity of the masking acoustic signal. Overcoming energetic masking primarily requires adequate separation of the acoustic features of the target from those of the masker (see Brungart (2001) for more information.) Success in separation depends on many factors, such as the masker's intensity and its temporal and spectral morphology (Bregman, 1990; Darwin, 2008).

Degradation without energetic masking occurs due to degradation of the source speech signal during transmission. This masking can be caused as a result of factors such as reverb, telephone line distortion and receiver limitations (as a results of hearing loss or hearing aid/cochlear implant processing) (Mattys et al., 2012). A person's ability to compensate for these distortions depends on many factors such as their degree/type of hearing loss and their ability to apply contextual cues to infer meaning from the degraded signal (Norris and McQueen, 2008).

Unlike the previously defined masking types, informational masking occurs due to its higher level contextual and semantic value, in addition to its acoustic properties. For example, this may occur in the presence of a competing talker where the content of the competing speech draw the listener's attention away from the target speaker. Informational masking is addressed by a listener's higher level cognitive processing such as attentional capacity and working memory (Kidd et al., 2008; Hoen et al., 2007). For an in-depth review of speech degradations and their effect on speech recognition performance, refer to Mattys et al. (2012) and Brungart (2001)

2.1.3 Effects of hearing loss on speech-in-noise performance

A hearing loss is typically categorised based on the location of the abnormality or lesion. Abnormalities or lesion at the middle ear are referred to as a conductive hearing loss, characterised as impaired conduction of sound signals through these areas. Damage to these systems can impact perception of speech, and may reduce speech recognition. As detailed in Section 2.1.1.4, access to binaural cues can impact a person's speech-in-noise performance. Damage to the outter ear may reduce access to these cues, such as deformation of the pinna or blockage of the ear canal as a result of excessive ear wax (Oldfield and Parker, 1984). However, steps can often be taken, such as wax removal or reconstructive surgery, which can have considerable positive impacts for addressing speech-in-noise performance issues related to a conductive hearing loss (Oshima et al., 2010; Rivolta, 2013). Some conductive losses even show improvements to speech-in-noise performance, an effect known as "paracusis willisii", caused by the attenuation of background noise by the conductive loss, and the increased intensity of the speaker's voice to account for the increase in background noise (known as the Lombard voice reflex, see Gelfand, 2016, p.139 for details). For these reasons, and due to the relative pervasiveness of sensorineural hearing losses (Carhart and Tillman, 1970; Pekkarinen et al., 1990; Yueh et al., 2003, p.278), this thesis will focus primarily on sensorineural impairments.

Sensorineural hearing losses are caused by neural pathway and cochlear pathologies. This type of hearing loss covers a broad range of issues, from impaired transduction from mechanical vibration to neural activity, to lesion in the neural auditory pathways. Retrocochlear related hearing losses (losses that occur as a result of pathologies beyond the cochlea) such as lesions of the eighth cranial nerve may have adverse effects on speech-in-noise performance. However, this section will focus primarily on lesions at the cochlea due to their relative prevalence.

Degradation of the cochlea and associated neural pathways due to factors such as high levels of noise exposure, ototoxicity, vascular issues and presbycusis can affect the functioning of the hair cells, neurons and other components that transduce mechanical vibration (Gelfand, 2016, p.137). This can result in distorted characteristics of the auditory filters, negatively impacting a listener's ability to discriminate spectral detail in speech (See Section 2.1.2.1 for details). For example, reduction in amplitude at high frequencies and impaired ability to discriminate frequencies are types of distortion that are particularly detrimental to speech-in-noise performance. These can be caused by conditions that affect outer hair cell function, which can result in the widening of auditory filter bands and increase linearity in terms of loudness perception (Glasberg and Moore, 1986; Oxenham and Bacon, 2003). These distortions to a patient's perception of spectral features and loudness can result in patients being able to detect, but not able to understand speech (Gelfand, 2016, p.138).

Current clinical interventions such as hearing aids and cochlear implants are regularly employed to treat sensorineural hearing losses. Success of these interventions vary, as discussed in Section 2.3.1. Alternative interventions such as electro-haptic stimulation may also offer improvements to speech-in-noise performance for those worst affected by hearing losses. These interventions are discussed in Section 2.4

2.2 Measures of speech-in-noise performance

An effective measure of speech-in-noise performance can offer insight into a patient's hearing ability for diagnostic purposes. It may also offer an indication of the performance of a patient's hearing interventions (such as a hearing aid or cochlear implant. See Sections 2.3 and 2.4 for details on traditional and alternative interventions). This measurement is required, both for diagnosis of impaired speech-in-noise performance and for development of better interventions to address speech-in-noise performance issues.

There are currently a wide variety of clinical measures for assessing speech-in-noise performance. These are performed behaviourally, by presenting a stimulus and assessing the response of the participant (responses are typically verbal or indicated by pressing a button). These tests are widely used, but may offer only limited information on the underlying mechanisms that drive speech-in-noise performance. The advantages and shortcomings of these tests are discussed in detail in Section 2.2.1.

An alternative method that is not currently used clinically for the assessment of speech-in-noise performance is the evoked response. These measures may offer an objective alternative to traditional metrics, and may also offer additional insight into the underlying neural mechanisms of speech-in-noise performance. A review of current evoked potential based methods are discussed in Section 2.2.2

2.2.1 Behavioural measures

Behavioural measures are the most commonly used measures in clinic for assessing both speech-in-noise performance, and for more general assessment of hearing. The most common types of measures are pure tone audiometry and speech-in-noise tests. PTA is a common clinical measure of general hearing performance. It is performed by playing tones of various audiometric frequencies to the patient and adjusting the level between presentations to assess the lowest level that the patient can detect. Patients respond by pressing a button when they hear a tone in either ear, until they can no longer hear a tone. For a full description of the PTA procedure, refer to Gelfand (2016) and British Society of Audiology (2018). PTA is the gold standard for assessing general hearing performance and is used for fitting and tuning of hearing aids and cochlear implants. It is suitable for clinic as it is fast to administer and provides intuitive, standardised results that can be used to categorise a hearing loss. However, it is suggested in many studies that PTA is a poor predictor of a person's speech recognition performance in noise (Killion and Niquette, 2000; Middelweerd et al., 1990; Carhart and Tillman, 1970). This is thought to be due to factors such as the influence of higher level cognitive processing, as well as integration of senses such as visual cues (and potentially vibrotactile cues, as discussed in Section 2.4) on speech recognition (Smoorenburg, 1992; Humes and Roberts, 1990), and the lack of representation for distortions caused by interfering noise sources (both externally as a result of additional background noise, and internally as a result of a hearing loss (Lee and Humes, 1993)). Plomp (1986) concludes that although speech in quiet scores may correlate with PTA derived measures, PTA measures may not correlate well with speech-in-noise scores. for speech in noise. Although PTA derived features such as the pure tone average correlate well with speech reception thresholds

(SRTs; The standard metric for assessing speech reception, as described in detail in Section 2.2.1.1) for attenuated speech, this is no longer true for SRTs for distorted speech or for SRT for both attenuated and distorted speech. This is thought to be due to the lack of representation for the effects of non-linear distortion in PTA results.

This is supported by studies conducted by Humes and Roberts (1990), Lee and Humes (1993) and Smoorenburg (1992). Humes and Roberts (1990) and Smoorenburg (1992), which consistently demonstrated that PTA scores do not account for a considerable amount of variance in speech recognition scores in noise, with R^2 values ranging from as little 0.3 to a maximum of 0.7 using various combinations of frequencies as predictors. Smoorenburg, 1992 illustrated the considerable inter-subject variability between participants when comparing SRTs with PTA scores. For example, when considering participants with SRTs at -2.5 dBA, PTA averages of 2 and 4 kHz range from around 0 to 55 dB HL. An exception which confounded previous findings is a study by Barrenas and Wikstrom (2000), which reports an R^2 of 0.92 when comparing PTA averages at 3, 4 and 6 kHz to speech recognition scores at a signal to noise ratio (SNR) of +4 dB. This study suggests that a non-linear relationship exists between PTA scores and speech recognition scores. However, methodology of the study limited the applicability of such findings to real-world performance (such as the use of only a single set SNR and use of speech shaped noise as opposed to noise that is more representative of everyday background noise. Further details on the implications of such factors is provided in Sections 2.2.1.2 and 2.2.1.3).

Many behavioural speech recognition tests have been developed in attempts to more accurately measure speech-in-noise performance. More specifically, tests that include the presentation of noise have been developed to better simulate the conditions that speech may be encountered in the real world. Factors such as the test format, speech stimulus and added noise all have a significant impact on the test's validity.

2.2.1.1 Speech-in-noise test procedures

Clinical speech and speech-in-noise recognition tests are typically behavioural measures. These are based on the assumption that a person's behaviour in a given condition forms an indirect measurement of their perception (Leek, 2001). Therefore, these measures aim to quantify a person's perception of speech based on their responses to a varying speech (and added noise) stimulus. Performance is measured as a function of speech intensity or SNR, with responses typically measured by analysing the person's ability to repeat the stimuli. These functions are commonly referred to as psychometric functions. This section aims to provide a brief review of the most prominent considerations when comparing speech-in-noise tests, and provides examples of some commonly used clinical tests. For a comprehensive overview of speech-in-noise/speech in quiet test methods, refer to Lawson and Peterson (2011) and Gelfand (2016).

There are three commonly used methods for sampling psychometric functions in order to provide a recognition score: by estimating a listener's SRT; by calculating percent correct responses to stimuli at a single intensity/SNR; or by calculating their performance across a range of equally spaced intensities/SNRs, known as the method of constant stimuli. Choice of method is determined by a number of factors, such as the type of information that is required from the test and the time allocated for the test.

One measure that provides information about the limits of a person's ability to understand speech is the SRT (Gelfand, 2016). The SRT estimates the point on a listener's psychometric function at which a listener can correctly recognise a certain percentage of presented words. This percentage is typically 50% (also known as the SRT50 or SNR50) but can be adjusted based on the method used for its calculation. There are a number of methods that can be used to calculate an SRT, most commonly using an adaptive track (refer to Lawson and Peterson (2011) for a comprehensive overview of methods for calculating SRTs). This procedure adapts the level of the signal and/or noise based on the percent correct score (the percentage of words correctly repeated out of all scored words in a sentence, for example) of a person's response. An advantage of this method is that, by varying the difficulty of trials based on previous responses, subsequent trials converge around the threshold efficiently, allowing for accurate threshold estimation with a minimum number of trials. An adaptive track is used in tests such as the hearing in noise test (HINT) in order to determine a listener's SNR based SRT (Nilsson et al., 1994).

The method of constant stimuli is an alternative method, which measures performance at equal intervals across intensities/SNRs, from imperceptible to consistently perceptible (Leek, 2001, p.1279). This method is used in tests such as the Bamford-Kowal-Bench (BKB) speech-in-noise (Niquette and Killion, 2016), QuickSIN (Killion et al., 2004) and Word in noise (WIN) test. This method allows for a better estimation of the function as a whole, but requires considerably more trials to gain the same level of accuracy around the SRT (typically calculated using the Spearman-Kärber equation — refer to Miller and Ulrich (2001) for details). One advantage of an accurate estimate across the psychometric function is the ability to accurately estimate its slope. An example is the use of a slope measure by Wilson et al. (2003) to demonstrate the shallower roll-off in performance of hearing impaired participants in comparison to normal hearing participants as the speech-in-noise test's difficulty increased. However, estimations of slope from adaptive track procedures significantly increase the number of trials needed, reducing the benefits in terms of speed of the procedure (Leek, 2001).

Several considerations with respect to the format of the test may impact the measured performance and can have a substantial effect on reliability of the test. These include: the size of the test, which forms a trade off between test duration and test reliability; the method listeners will use to respond, affecting the influence of chance correct responses based on the open or closed set format of the test; and the scoring method, which affects the relative influence of lower and higher level processing on a participant's score. For brevity, the reader is referred to Gelfand, 2016, p.215-242 for a comprehensive overview of the effects of these parameters on speech-in-noise tests.

2.2.1.2 Speech stimulus

Stimuli vary considerably between the most prominent speech materials used for speech-in-noise/recognition tests, with a range of different characteristics that influence the quality of the metrics. Stimulus types range from individual syllables and isolated monosyllabic or spondaic words to full sentences, with aspects such as linguistic complexity, semantic context and acoustic content varying considerably. Although the exact relationship between cognition and speech perception is not fully understood, an awareness of the effects that stimulus choice has on the underlying processing is crucial, as it determines the auditory functions that will be measured by any recognition/speech-in-noise test. Lawson and Peterson (2011) define speaker variability, stimulus type and word familiarity as key factors that affect speech recognition performance. This section outlines current understanding of the relationship between speech stimuli selection, recognition performance and the processing mechanisms that contribute to such performance.

The type of stimuli chosen for a recognition/speech-in-noise test primarily affect the influence of higher level cognitive abilities influence test results. The types of cognitive abilities that contribute to these tests include working memory capacity, attentional capacity, speed of linguistic processing, knowledge of language vocabulary, and ability to make contextual inferences (Grant and Seitz, 2000; Moradi et al., 2014; Humes et al., 2013; Humes and Dubno, 2010). For example, performance on speech-in-noise/speech in quiet tests have been shown to be influenced increasingly by working memory and attentional capacity as a function of stimulus length (Moradi et al., 2014; McArdle et al., 2005). In the case of sentences, a reduced ability to focus on and recall multiple words after a stimulus is presented may negatively impact recognition scores.

A participant's ability to infer information from contextual cues has also been shown to have significant impact on scores, particularly when using sentences that have substantial semantic and syntactic context. These cues allow listeners to use knowledge of the language and information surrounding a word to surmise a reasonable answer when auditory cues are lacking (Ernestus et al., 2002; Mattys et al., 2012). Use of isolated syllables or words in speech recognition tasks have been shown to minimise the contextual information available to a participant (Lawson and Peterson, 2011). This places focus on the participant's ability to process lower level acoustic features, limiting the level to which these top-down influences, such as attention and working memory, affect results. Ernestus et al. (2002) demonstrate the effect that context has on reduced speech using Dutch words surrounded by various degrees of contextual information. Results suggest an increasing reliance on surrounding context as a target word's acoustic content is reduced. The stimulus types used for a range of speech materials is detailed in Table 2.1

It is well known that vocal qualities vary from person to person. Factors such as gender, age, accent and articulation may affect performance and should be considered when choosing material for a speech-in-noise/recognition test. These factors are thought to impact speech recognition performance due to the differences in acoustic characteristics between male and female, as well as young and old speakers. Aspects that differ include the fundamental frequency (f_0), overall intensity and more complex characteristics, such as 'breathiness', which are discerned through a variety of acoustic cues (Klatt and Klatt, 1990; Lee et al., 1999).

Variation in speaker accent (both non-native and unfamiliar native) may also present significant challenge to a listener, and has been linked to reduction in language processing speed and increased listening effort. In the case of heavily accented speech it may result in complete lack of comprehension (Adank et al., 2009; Anderson-Hsieh and Koehler, 1988). The divergence from clear speech caused by speech disorders and disfluencies common in everyday conversation can also have a negative impact on speech intelligibility. A detailed review of the effects of clear speech on a listener's performance is discussed by Smiljanić and Bradlow (2009). A comprehensive analysis of the many variations in speaker characteristics is outside the scope of this review, however further insight is provided by Uchanski (2005) and Mattys et al. (2012). These factors will primarily affect the level to which a listener will need to rely on higher level compensatory strategies, placing pressure on cognitive functions such as working memory, attentional capacity and lexical mapping. The speaker type of the chosen materials will largely depend on the source of the recordings. Variations in recordings of materials may affect a listener's performance and therefore care should be taken when choosing appropriate stimuli.

Material name	Stimuli	Test response/scoring	No. list items	No. lists	Set type	Source
PAL PB-50	Monosyllabic words	Repeat word	50	20	Open	
CID W-22	Monosyllabic words	Repeat word	50	4	Open	
NU-6	Monosyllabic CNC words	Repeat word	50	4	Open	
CID Everyday Sentences	Sentences	Repeat sentence	10 (50 key words)	10	Open	
NST	High-frequency nonsense syllables	Repeat syllables	25	12	Open	
MRHT	Monosyllabic words	Repeat & mark the word (1 of 6 alternatives)	50	4	Closed	
CCT	High-frequency monosyllabic words	Check the word (1 of 4 alternatives)	100 & 50	4	Closed	
SSI-ICM & SSI-CCM	Synthetic sentences	Identify sentence by number (1 of 10 alternatives)	10	24	Closed	
CUNY NST	Nonsense syllables	Mark the syllable	7–9	11 subtests	Closed	
IEEE sentences	Phonetically balanced sentences	—	10	72	—	
BKB sentences	Sentences	Scored based on key word recognition	21	50	—	Bench et al. (1979)
AB words	Sentences	Scored based on key word recognition	21	50	—	Bench et al. (1979)

Table 2.1: Summary of speech recognition test materials

2.2.1.3 Noise stimulus

Smoorenburg (1992) demonstrated considerable variance in the relationship between a listener's speech-in-noise performance and audiogram. This is thought to be as a result of this test measuring primarily the listener's ability to withstand the effects of attenuation, with no explicit consideration for the influence of signal degradation (beyond any further attenuation of a signal that may occur as the result of a degradation). The addition of distortion is therefore necessary to gain a more in depth understanding of a listener's real-world speech recognition performance. The choice of noise stimulus/distortion determines the types of real-world speech degradations that may be encountered by a listener. There are 3 types of degradation: Energetic masking, non-energetic masking and informational masking. Energetic masking occurs as a result of interference from another source (background noise), non-energetic masking occurs as a result of source signal degradation (distortion as a result of hearing aid processing, from degradation as a result of a hearing loss or via a telephone line or for example) and informational masking occurs as a result of interfering higher level information from a competing speech source (Hoen et al., 2007). As energetic and informational masking are consistently reported to be a significant issue for hearing aid users and for those who suffer from hearing loss, they should be primary considerations when developing/administering speech-in-noise tests (Dubno et al., 1984). This section will therefore focus on these forms of masking, outlining the types of noise/distortions that are typically used for speech-in-noise tests. The implications of noise type selection and presentation method on a listener's speech-in-noise test performance will also be discussed.

When considering the effects of energetic masking, the relationship between the spectral and temporal characteristics of the speech signal and the noise that play a significant role in the listener's ability to recognise the speech (Theunissen et al., 2009; Dreschler et al., 2001; Soli, 2008). The two most common types of noise used in speech-in-noise tests are speech shaped noise (white noise shaped to the long term spectral characteristics of speech) and babble noise (a combination of a number of overlapping background speakers), each of which have varying spectral and temporal characteristics.

Babble noise has been shown to be an effective masker for speech-in-noise tests and has been used in tests such as QuickSIN (Killion et al., 2004), WIN (Wilson and Burks, 2005) and BKB-SIN (Niquette and Killion, 2016). Babble noise is generated by mixing various numbers of recorded speakers and is thought to be a type of noise with particular ecological validity, emulating an environment where a listener must recognise target speech amongst various numbers of competing speakers (Cullington and Zeng, 2008; Wilson et al., 2003). Unlike speech shaped noise, babble noise contains spectral and temporal properties that vary over time. Most notably, spectral and temporal 'dips' have been shown to contribute to a listener's performance. These dips arise in the temporal domain from short pauses in the masking speech, and at points where the masking speech has low energy — typically at unvoiced consonants such as ‘m’, ‘n’, ‘k’, or ‘p’ (Peters et al., 1998). In the spectral domain, dips arise due to the varying spectral qualities of the masking speech over time, and the relative variations in the target speech. When the masking speech does not fully overlap the target speech in the spectral domain, it may be possible for a listener to recognise unmasked portions of the target speech. Choices such as age, gender and number of speakers used to generate the babble noise have significant impact on temporal and spectral properties of the noise (Duquesnoy, 1983; Summers and Molis, 2004; Drullman and Bronkhorst, 2004; Hawley et al., 1999; Middelweerd et al., 1990). The number of speakers used to generate the noise is of particular importance with regards to energetic masking. It has widely been observed that as the number of competing speakers increases, temporal and spectral fluctuations decrease (Bronkhorst and Plomp, 1992; Simpson and Cooke, 2005; Drullman and Bronkhorst, 2000; Hawley et al., 1999). This is caused as dips are reduced when further speakers are added (Bronkhorst and Plomp, 1992). However, discrepancies in the effects of speaker number between studies have highlighted the dependence of this variable on other factors that may affect the acoustic content of the babble noise, such as speaker gender, accent, pitch etc. (Rosen et al., 2013; Bronkhorst, 2000; Cherry, 1953).

Alternatively, speech shaped noise has been used in tests such as HINT (Nilsson et al., 1994) and for improvement of SRT estimation methods by Plomp and Mimpen (1979). Speech shaped noise is generated by filtering white noise, typically by generating the long term average speech spectrum (LTASS) of the target stimulus. The resulting noise has a constant energy at frequencies approximately equal to the energy at those frequencies in the speech stimulus. The steady temporal and spectral characteristics of such noise ensures an approximately equal SNR across frequencies, regardless of speaker variabilities such as gender (Nilsson et al., 1994; Theunissen et al., 2009). However, as steady state speech shaped noise is constant in both the spectral and temporal domains, it fails to account for differences in listener performance as a result of spectral or temporal dips. Studies often modulate the temporal characteristics of steady noise in order to produce a temporally modulated noise that maintains its static spectral properties (Schoof and Rosen, 2014; Buss et al., 2009; Middelweerd et al., 1990). This emulates the temporal dips commonly found in natural noises such as babble noise, whilst continuing to fully mask the target speech in the spectral domain. This type of noise has been used in speech-in-noise tests such as the Matrix test (Hagerman, 1982).

A key difference between the presented noise types is the lack of informational masking when using steady state and amplitude modulated noise (Theunissen et al., 2009). As babble noise is comprised of various streams of intelligible speech, the recognition of speech from competing speakers may interfere with target speech comprehension during higher-level lexical processing. This has been demonstrated in a variety of studies, showing that informational masking may interfere despite the lack of energetic masking (Spieth et al., 1954; Simpson and Cooke, 2005; Sperry et al., n.d.; Van Engen and

Bradlow, 2007) and that when interference is a competing speaker, informational masking may be the dominant cause of interference (Brungart et al., 2006).

It is clear that choice of noise type affects the types of auditory processing and the levels to which higher-level cognitive ability contribute to a listener's performance in a speech-in-noise test. A study by Wilson et al. (2007) compared the relative effects of babble and speech shaped noise on speech-in-noise test results. Results suggest that normal-hearing listeners tend to perform better by around 2.1 dB in babble noise, with only minor differences in psychometric function morphology. The intrasubject standard deviation of normal-hearing listener's SRTs were 1.0 dB for speech shaped noise and 1.3 dB for babble noise (Wilson et al., 2007). These results are supported by similar results presented by Plomp and Mimpen, 1979 and those presented by Wagener and Brand, 2005. Wilson et al., 2007 also reported increased standard deviations for hearing-impaired listeners — 3.5 dB for speech shaped noise and 4.1 dB for babble noise. Overall, results suggest that babble noise may increase test-retest variability when compared to speech shaped noise. However, the increase in ecological validity due to the addition of informational masking offers a significant advantage when considering a speech-in-noise tests applicability to performance in a real-world situation, such as listening to speech amongst competing speakers.

2.2.1.4 Common clinical tests

To illustrate the parameters outlined in the previous Sections 2.2.1.1 to 2.2.1.3, this section outlines examples of tests used in The University of Southampton's Auditory Implant Service (USAIS) for assessment of cochlear implant users and candidates. It should however be noted that testing procedures may vary accross centres, based on factors such as location and the clinical populations that they attend to. Table 2.2 outlines the parameters of five speech-in-noise tests used at USAIS.

2.2.2 Electrophysiological measures

Auditory evoked responses are measurements of synchronous electrical activity in the auditory system in response to sounds. By placing transducers on a listener's head, it is possible to monitor the change in the auditory nervous systems' field potentials in response to a stimulus. This has been shown to provide valuable insight into the functioning of the auditory processing system for both research and diagnostic purposes (Gelfand, 2016). Auditory evoked responses are typically measured using EEG, MEG or ECochG. These neuroimaging techniques have a number of attributes which make them particularly well suited to analysis of the auditory system. In particular this includes their high temporal resolution, the objective nature of these measurements and

Test name	Parameter	Value
BKB Adaptive Test for adults	Speech material	BKB Male
	Lists used for each session	Software selects trials at random across all BKB Male lists
	Noise material	Speech-shaped noise
	Adaptive track	2 down, 1 up
	Adapted parameter	Noise
	Speech level	65 dBA
	Starting SNR	+10 dB SNR
	SNR calculation	RMS of speech and noise material calculated (removing silent sections), then ratio calculated
	Step size	5 dB, 1 dB
	Number of reversals	2 big, 8 small
	Scoring	SNR — The average of the last 8 reversals
	Transducer	Mono central speaker
BKB Non-adaptive In-quiet Test for adults	Speech material	BKB Male
	Lists used for each session	Consecutive lists, noting previously completed lists to avoid repetition
	Speech level	65 dBA
	Scoring	Percentage key-words correct for 2 lists
	Transducer	Mono central speaker
Automated Toy Test for children	Speech material	Automated Toy Test sentences
	Noise material	Speech-shaped noise
	Adaptive track	2 down, 1 up
	Adapted parameter	Speech
	Noise level	55 dBA
	Starting speech level	+15 dB SNR
	Transducer	Central speaker, 90°left and 90°right speakers
	Speech location	Centre
	Noise location	Centre, 90°left, 90°right
	Step size	8 dB, 4 dB
	Number of reversals	2 big, 8 small
	Scoring	SNR — Average of last 8 reversals, then subtract the noise level
AB Words Non-adaptive Test for adults	Speech materials	AB Words lists
	Speech level	65 dBA
	Scoring	Complete words correct and syllables correct
	Transducer	Central speaker

Table 2.2: Summary of speech recognition and speech-in-noise tests used by USAIS

the lack of need for a response from a listener, making them suitable for assessment of patients that are unable to respond to behavioural measures.

Electrophysiological measures also show potential for research purposes, as a method for analysing the underlying neural mechanisms that drive speech-in-noise performance. This could provide further insight into the auditory processing mechanisms as well as being extendable to analyse the integration of other senses that result in improved speech-in-noise performance (Walter, 1964; Bourguignon et al., 2020; Riecke et al., 2019). This thesis will focus on auditory evoked responses, however discussion is provided in Section 7 on the possible use of multisensory evoked responses in future work.

There are also a number of issues that must be addressed when considering evoked responses, including poor SNR in the measured responses, poor spatial localisation of the generator of the response (in comparison to the accuracy of methods such as fMRI of fNIRS, for example) and the range of practical limitations that are inherent in these methods. For brevity, this review will focus primarily on measurement using EEG, due to its wide availability, relatively low cost and non-invasive nature. MEG based literature will also be considered due to its analogous response with those recorded using EEG. This section will provide an overview of the different forms of auditory evoked responses and the methods used for obtaining such measurements. Relevant research into their applicability for the assessment of speech-in-noise performance will also be discussed.

Evoked responses using EEG measure the field potential recorded via electrodes, placed across the participants scalp, in response to a stimulus. These potentials are produced as the summation of extracellular electrical activity from populations of neurons in the central nervous system. When recorded in response to an auditory stimulus, this produces an auditory evoked response. EEG is classified as a far-field recording technique, as measurements are taken at considerable distance from the source of the electrical activity (in comparison to near-field techniques, such as EcochG, where measurement is taken as close to the source of activity as possible). Therefore, the number and positioning of electrodes has a substantial impact on the recording quality. Auditory evoked responses are recorded using between 3 and 128 electrodes which are placed across the scalp in various configurations. The number of electrodes used and their positioning are commonly dictated by a predefined schema known as the 10–20 system, which details the number and placement of electrodes (Libenson, 2010). The choice of electrode montage can affect the spatial resolution of the resulting recording, as a montage with fewer electrodes spaced at greater relative distances will not be able to resolve the location of neuronal activity accurately as a tightly spaced montage with many electrodes (Niedermeyer and Lopes da Silva, 2005).

The underlying spread of current is also a limiting factor in spatial resolution, as increasing the number of electrodes decreases the independence of each electrode. This is due to the electrical current produced from neuronal generators reaching multiple electrodes.

For brainstem response, as few as 3 electrodes are sufficient to measure a response, however use of such a reduced number of electrodes is primarily for practical purposes in a clinical environment, particularly in reducing setup time (Hall, 1992). When considering responses such as cortical responses in a research context, a larger number of electrodes are typically used. It should also be noted that activity is measured relative to a reference. This may be a single reference electrode for all electrodes or may vary depending on the chosen montage. Using a reference electrode with a differential amplifier allows for the high levels of external electrical noise that are common to both electrodes to be rejected (known as common mode rejection), leaving primarily the relatively low voltage action potentials to be measured in the output signal (Hall, 1992; Libenson, 2010). Having been measured and amplified the signal is finally converted from analog to digital using and analog-to-digital converter ready for further digital processing.

A particular issue with EEG is the attenuation of neural activity measured at far field (at the scalp) and the interference from multiple neural generators. Due to the many layers of tissue that must be traversed for a potential to reach the scalp, a large number of neurons must be activated simultaneously to achieve a measurable response. MEG is a comparable alternative form of encephalography that suffers less from these issues, however this method requires a considerably more controlled environment and is often not available or practical for clinical purposes.

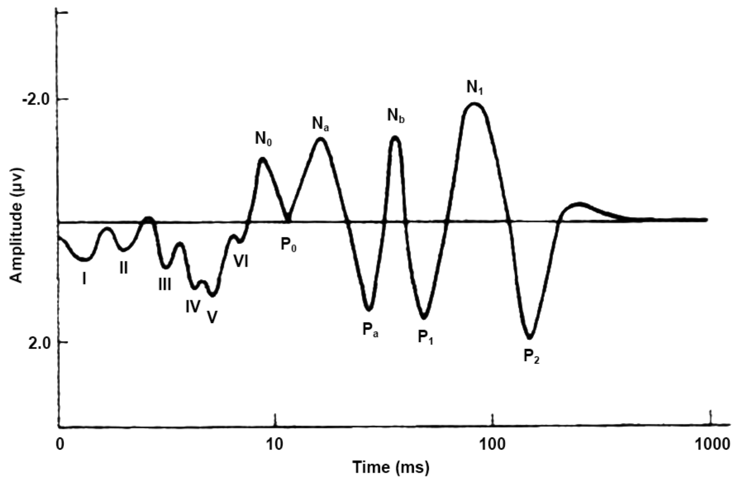

Figure 2.3: Example of an idealised response to click stimulus[3]

There are a number of considerations which can impact the quality of evoked responses. The most significant factors are outlined by Hall (1992):

- Analysis time (typically the number of epochs measured)

[3]This figure was adapted from Albert Kok—Eigen bewerking van Picton et al. (1974)., Public Domain, `https://commons.wikimedia.org/w/index.php?curid=2213490` (accessed 07/12/2021)

- Noise
- Artefacts
- Filter settings

The size of epochs measured detemines the time period after the onset of a stimulus that is analysed. This has implications for the analysis of underlying sources in the central auditory pathways, as it is understood that different components respond to a stimulus at various latencies. Figure 2.3 illustrates the signal morphology of the prototypical click responses: The auditory brainstem response (ABR), auditory mid-latency response (AMLR), and auditory late response (ALR). Table 2.3 provides details on some commonly measured responses to a click stimulus, discussed in Section 2.2.2.1 (with the exception of the auditory steady state response (ASSR), discussed in Section 2.2.2.2).

Type	Latency (msec)	Terminology	Presumed source/generator
Early	<12		
EcochG	1–4		
		Cochlear Microphonic	AC component receptor potential—cochlear hair cells
		Summating Microphonic	DC component receptor potential—cochlear hair cells
		N_1 or AP	Auditory nerve compound action potential I-auditory nerve action potential
ABR	1–12		Brainstem
		Wave I	Compound action potential recorded from the distal end of the acoustic nerve or graded potential of dendritic terminals of acoustic nerve
		Wave II	Changes in current flow at the poms acusticus internus, or compound action potential of the auditory nerve at the entrance into the brainstem, or graded potentials from cochlear nucleus
		Wave III	Cochlear nucleus and trapezoid body or superior olivary complex and trapezoid body
		Wave IV	Lateral lemniscus, ventral lemniscus cells, or superior olivary complex or ascending auditory fibers in the pons
		Wave V	Ventrolateral inferior colliculus and ventral lateral lemniscus
		Wave VI, VII	Higher brainstem structures such as the medial geniculate body
Mid-latency	12–50		
Transient		Na (N20), Pa (P30)	Myogenic vs. neurogenic source
Late response	>50		
Late	50–250	N100, P150, N200	Cortical
Long	>250	P300, CNV	Cortical
Steady-state	Frequency domain response	40 Hz ASSR	Cortical
		80 Hz ASSR	Brainstem + cortical

Table 2.3: Summary of commonly measured evoked responses to auditory stimulus[4]

Given the extremely small amplitude of the signals being measured in evoked responses, one of the most significant issues is how to account for the noise generated by other sources of activity, both internally and externally. For example, the amplitude of a wave V may be as little as 0.5 μV, which is orders of magnitude smaller than the combined noise from internal sources such as other brain activity, muscle activity and external sources such as electrical activity from power lines (Hall, 1992). Common mode rejection is commonly used to account for this poor SNR. As EEG measures a potential between two electrodes, common mode rejection inverts the polarity of one electrode (the inverting electrode) and adds the resulting signals to the non-inverting electrode. This results in any signal common to both electrodes being subtracted from the final signal. Assuming that the noise to be removed is constant across both electrodes, this method significantly reduces the amount of background noise (typically by around 80 dB (Hall, 1992, p.69)), but does not have a significant effect on the response which should only be present in the non-inverting electrode's signal.

Another common process is to repeat the stimulus multiple times and average the resulting signals. This method of noise reduction assumes that the underlying response will remain constant across repeated presentations, allowing the stochastic noise to be averaged out. This has been shown to be effective for short stimuli such as clicks and tone pips and has traditionally been used in conventional evoked response measurements. However, recent methods have been developed that do not require multiple repeats of a stimulus to obtain a response. These methods have been shown to be effective in the analysis of speech evoked responses, as discussed in Section 2.2.2.5.

Specific artefacts in the signal can also be detected and attenuated using filtering methods. Three common examples are the noise generated by eye blinks, breathing and powerline electrical activity. Digital filters are commonly applied to detect and attenuate the effects of such artefacts. In the case of electrical activity for example, a narrow notch filter at 50 or 60 Hz (dependant on country) may attenuate such activity without significantly affecting the remaining signal. Effective filtering methods have been developed to attenuate eye blinks as a post-processing stage using independent component analysis. The details of the machine-learning based filtering approach used to achieve this are complex and outside the scope of this review (refer to Hoffmann and Falkenstein (2008)), however it has been shown to be an effective method for reducing the effect of such artefacts.

Traditionally, evoked responses have been recorded using short stimuli such as clicks and tone pips, repeated multiple times to elicit a response. This allows for these individual responses to be averaged, thus reducing noise across the recordings to reveal the consistent underlying response across repetitions. This technique is used in the ABR, AMLR and ALR. Alternatively, responses can be measured from a constant modulating stimulus, as is the case with the ASSR. These techniques each provide information

[4]This table was adapted from Niedermeyer and Lopes da Silva (2005, p.976, 980)

about auditory processing at various levels of the central auditory pathways, although their relationship to perception of speech-in-noise is not entirely clear (as discussed further in the following sections). In addition, it has recently been shown that speech and speech like stimulus can be effective for the analysis of the auditory system's processing of speech using evoked responses. These stimuli have the advantage of better representing real life speech than synthetic tones and clicks (Kraus and Nicol, 2003; Russo et al., 2004; Johnson et al., 2005). Types of stimulus range from analysis of responses to individual phonemes (such as /da/) to analysis of running speech. This section will first review current understanding of conventional stimulus, such as clicks and tones, including the presentation of such stimulus in noise, and discuss the relationship between these responses and speech-in-noise performance. The most prominent research into responses using speech based stimulus will then be reviewed, including the range of signal processing techniques that allow for the use of such stimuli.

2.2.2.1 Auditory brainstem response (ABR)

The ABR has been shown to be an effective diagnostic tool for clinical applications such as infant screening, threshold estimation and identification of pathologies in the auditory system. Click evoked responses are currently the most widely researched type of response, although tone bursts and chirps are also used (Hall, 1992, p.307-308).

Analysis of the brainstem response is performed in the time domain and focuses on the relative latency and amplitude of peaks (labelled I-V in Figure 2.3, see Table 2.3 for the physiological sources of these peaks) in the first 12 milliseconds after the onset of a stimuli. Alternatively, chirps (rapidly sweeping tones) and tone pips have been shown to effectively elicit responses. These stimuli allow for raised peak amplitude and frequency specific responses. These techniques are most commonly used for improving wave V amplitude and for the estimation of frequency specific thresholds of infants, as the click-ABR is thought to produce relatively inaccurate threshold estimations, particularly at low frequencies (Hall, 1992, p.424).

A small number of studies have suggested that the click-ABR can provide a predictor for speech discrimination performance under certain conditions. Borg (1982) and Rawool (1989) reported significant correlations between speech in quiet scores and various wave morphology features for listeners with acoustic neuromas ($N = 12$) and older normal-hearing listeners ($N = 7$). However, Borg (1982) found no correlation between speech discrimination and ABRs for sufferers of cochlear hearing loss. This is further supported by the finding of Makhdoum et al. (1998) in cochlear implant users ($N = 15$). It is also thought that responses to simplistic stimuli may not provide robust approximations of complex sounds such as speech (Greenberg, 2004; Song et al., 2006; Johnson et al., 2008, p.169-170) and may therefore be poor predictors of a listener's speech processing or speech-in-noise performance in general. It is therefore expected that ABR is unable to fully predict speech and speech-in-noise performance due to the lack of higher level cognitive processing involved in the neural responses to click and similar simplistic stimuli in the auditory system (Song et al., 2006). The focus of ABR on only lower level linear processing structures in the brainstem will only provide insight into the function of a limited subsection of the structures involved in speech processing. For these reasons, short speech based stimuli have also been used to elicit responses analogous to ABR but with greater ecological validity. These speech based stimuli are discussed further in Section 2.2.2.4.

Research into the effects of broadband noise on click ABRs have shown changes in ABR morphologies for various demographics as a result of increasing noise levels. A series of studies found various effects on response morphology as a result of increasing broadband noise level for both normal-hearing and hearing impaired listeners. Effects include increases in wave V latency, wave I-V interval and decreases in average wave amplitude (Burkard and Hecox, 1983b; Burkard and Hecox, 1983a). However, to the authors knowledge there is no research suggesting that these correlate with SIN performance.

2.2.2.2 Auditory steady state response (ASSR)

Modulated continuous stimuli using pure tones are used to produce the ASSR. As with tone pip elicited ABR, the ASSR can be used to determine the brainstem response to specific frequencies. By modulating the amplitude (and sometimes frequency) of a carrier wave and performing spectral analysis of the resulting response, a peak at the modulation frequency can be used to determine the brainstem response to the carrier frequency. As the ASSR is elicited by a continuous stimulus, overlapping windows are analysed in the frequency domain and averaged to produce a response that is a summation of synchronous activity from sources throughout the central auditory pathways (Picton et al., 2003). The relative contributions of these sources are determined primarily by the rate of stimulus amplitude and frequency modulation. Stimuli are typically modulated at either 40 Hz, eliciting responses primarily from both the brainstem and primary auditory cortex, or at 80–100 Hz, where responses are produced primarily in the brainstem (Herdman et al., 2002).

As with ABR methods, the simplistic nature of the ASSR stimulus may limit its applicability to the prediction of speech-in-noise performance in terms of measuring higher cognitive functioning. However, studies such as that by Dimitrijevic et al. (2004) (using young normal-hearing listeners $N = 10$, elderly normal-hearing listeners $N = 10$ and elderly hearing-impaired listener $N = 10$) and Dimitrijevic et al. (2001) (using young normal-hearing listeners $N = 21$) report significant correlations (ranging from between $r = 0.38$ and $r = 0.74$) between independent amplitude and frequency modulated ASSRs and word recognition scores in quiet. These correlations are attributed to the similarities between modulations in the ASSRs and those found in speech. Significant correlations are also thought to be produced as a result of the combined contributions of processing at the various stages throughout the central auditory pathways. A study by Manju et al. (2014) reported correlation ($r = 0.61, p < 0.05$) between SRTs obtained using an adaptive speech-in-noise test (which comprised on HINT derived materials and procedure, adapted for indian speech), and 30–40 Hz ASSRs for normal-hearing listeners ($N = 11$). However it is noted that predictions were not possible at higher modulation frequency ASSRs. Therefore an ASSR that stimulates responses at the brainstem level may not provide a similar level of predictive performance. It should also be highlighted that group level correlation with speech-in-noise scores does not directly suggest that a measure could be developed to accurately predict scores at the individual level. Further work would be needed to ascertain within-subject variability.

Recent research has also explored the potential for speech spectrum ASSRs presented in noise to be used for speech-in-noise performance. In addition to their work using ASSRs in quiet, Dimitrijevic et al. (2004) compared ASSRs with speech spectrum noise at various modulation rates to word-in-quiet scores, finding that both young and elderly subjects had similar reductions in response recognitions and amplitudes when ASSRs were presented in noise as opposed to in quiet. Leigh-Paffenroth and Murnane (2011) found significant correlations between WIN scores and the 40 Hz ASSR in babble noise when measuring ASSR detection rate and average ASSR amplitude in normal-hearing listeners ($r = -0.38, p = 0.044$) and HIs ($r = -0.39, p = 0.039$). However, although significant, these results accounted for only 15% of the WIN score variability for normal-hearing listeners and only 23% for HIs. The large inter-subject variability suggests that this measure may not be a robust predictor of behavioural speech-in-noise measures. Overall, the presented research suggests that ASSR provide robust correlates for word recognition in quiet scores at the group level. However, as these scores do not account for the distortion factor described in Section 2.2.1.3, these may not translate to similar predictive capabilities for speech-in-noise tests. There has been considerably less success in correlating ASSRs, recorded both with and without added noise, to speech-in-noise scores.

2.2.2.3 Mid and late latency responses (AMLR and ALR)

Using a similar method to that for recording an ABR, the AMLR and ALR can be recorded to analyse synchronous activity at higher levels of the central auditory pathways. These measures provide responses to cortical and mid-level structures that are not represented in ABR. Refer to Figure 2.3 for illustration of the AMLR and ALR response morphology.

The AMLR occurs at 12–50 ms and is comprised of four waves: Pa, Na, Pb, Nb (P=Positive, N=Negative). These waves are thought to occur as a result of activity in mid and higher level structures such as the inferior colliculus and primary and secondary structures in the auditory cortex (Hall, 1992, p.553-558). Few notable studies exist that directly measure the relationship between the AMLR and speech recognition or speech-in-noise. Romero et al. (2015) found weak correlations for Na-Pa amplitude and Na latency with syllabic awareness test for children with learning disabilities. However, studies by Paludetti et al. (1991) and Makhdoum et al. (1998) found no correlation between AMLR and speech recognition performance in 74 normal-hearing listeners and 15 cochlear implant users. It was found that AMLR morphology could often appear normal alongside abnormal speech recognition results and vice versa. This suggests that there may be some link between AMLR and higher level lexical processing. This research suggests that it is unlikely that the AMLR will provide a robust correlate for speech-in-noise performance.

The ALR occurs at latencies of more than 50 ms and primarily consists of components: P1 (thought to be equivalent to the Pb peak of the AMLR), N1, P2, N2, P300 and the mismatch negativity response (MMN). ALRs are typically elicited with tones or speech of significantly longer duration than those used for ABR or AMLR. Unlike ABR and AMLR, a wide variety of stimuli can be used for effective production of the ALR including the use of music and stimulus in noise. A further difference in the case of the P300 and MMN is that these components are elicited in response to change in stimulus, requiring a target stimulus to be presented infrequently amongst another frequent stimulus. These responses therefore provide an indicator of stimulus discrimination in neural encoding of a sound, as opposed to stimulus presence as is the case for other responses discussed previously (Hall, 1992, p.643–644, 716, 769). In terms of speech performance evaluation, the ALR is typically evoked in response to short speech stimulus. Research has shown correlation between the P300 and P2 components of the ALR with speech recognition performance in cochlear implant users (Makhdoum et al., 1998; Groenen et al., 2001). The ALR is thought to be more effective for estimation of speech-recognition performance used for cochlear implant users in these studies due to the contamination of lower level responses such as ABR by the electrical activity of the cochlear implant. However it should be considered that the measurement of responses such as the P300 may be considerably more difficult due to the need for active attention, again limiting the applicability of such techniques in hard to test patients and infants. Research into the use of ALR for prediction of speech-in-noise performance using short speech stimuli is discussed in Section 2.2.2.4.

2.2.2.4 Phonemes and words

A range of stimuli have been used for eliciting complex stimulus ABRs (cABR) for analysis of the brainstem response to speech-like stimulus. The most common stimulus is the consonant-vowel syllable /da/, although many other stimuli, both speech and non-speech based, have been used. cABRs are typically recorded with parameters analogous to conventional click-ABRs. As with conventional click-ABR, a range of defining features have been outlined in the literature, based on peak amplitude and latency, as well as features such as the frequency following response (FFR), which is determined by the f_0 of the steady state section of the stimulus. Responses to cABRs are averaged in order to improve SNR and so the limitation of stimulus length is still imposed on responses. For this reason, it is suggested that a stimulus should be between 40–100 ms to allow for reasonable recording times, although longer durations may be used (Skoe and Kraus, 2010). As with conventional ABR methods, the need for averaging cABRs may limit the ecological validity of the stimulus. For example, Reichenbach et al. (2016) and Neupane et al. (2014) suggest that repetitions of short stimuli may result in neural adaption which limits the applicability of such responses to the task of better understanding speech comprehension. This method has nonetheless been the focus of much research into the encoding of speech in subcortical neural pathways. Examples of such research include demonstrating that a robust representation of speech is maintained at the brainstem (Anderson et al., 2013) and that the encoding of the speech envelope is most important when listening to clean speech, whereas the temporal fine structure is increasingly exploited as SNR increases for speech-in-noise (Bidelman, 2016).

Multiple studies have explored the relationship between speech-in-noise performance and cABR responses. A series of studies by Anderson et al. (2010b), Anderson et al. (2010a), Anderson and Kraus (2010) and Anderson et al. (2013) found a number of correlates to HINT and speech, spatial and quality of hearing scale (SSQ) scores. In analysis of brainstem responses by Anderson et al. (2010b), it was found that f_0 responses in the consonant-to-vowel transition period of the stimulus correlated with HINT ($r = -0.424, p = 0.008$) scores. Anderson et al. (2013) also reported that the wave morphology variables they selected provided additional predictive power in SSQ scores to that provided by QuickSIN scores+participant age+standard click ABRs when analysing using multivariate linear regression (a significant change in $R^2 = 0.158, F(3, 103) = 5.413, p = 0.001$). It was noted that f_0 representation in the FFR did not significantly correlate with HINT or QuickSIN scores. These results contradict a considerable amount of literature (as outlined in Section 2.2.2.1) that suggest ABR based measures do not well predict speech-in-noise performance, by presenting evidence that additional predictive value may be provided by ABR measures evoked using speech-based stimulus. The authors highlight evidence that the cABR provides heightened sensitivity to “subtle differences” in hearing-impaired populations, citing Song et al. (2006), which may provide some explanation as to this discrepancy, however the exact cause is not clear and an area that would require further research. Although these results suggest that correlates to speech-in-noise performance metrics can be extracted from cABR, the intersubject variability is high and any one of these correlates may not reliably predict speech-in-noise performance alone. Furthermore, work by Novis and Bell (2019) suggests that cABRs may require further epoching than that of standard click stimulus to achieve reliable responses. This may limit its clinical applicability due to the increased time required for testing. Further research is required in these areas to independently verify the results found by Anderson et. al, and to determine the cause of the discrepancy between these results and those of ABRs to simple stimuli.

Research by Billings et al. (2013) also used a short speech like stimulus (/ba/) to study the effects of noise on ALR morphology. In addition, these results were compared with the participants ($N = 15$) SRTs, measured using IEEE sentences in speech spectrum noise. Results suggested that N1, P2, and N2 latencies decreased and amplitudes increased as a function of SNR. It was noted that that changes in ALR morphology due to the rollover effect (where speech signals become less intelligible at high signal levels) were not observed in this study. This may suggest that degradation in speech performance as a result of this effect may not be represented in this type of response. However it should be considered that the participant sample size was small and only included young normal-hearing listeners, therefore these results may not generalise to wider populations.

Single words have also been used by epoching at set syllables and averaging responses. This technique was used by Wagner et al. (2016), with the aim of understanding the relationship between P1-N1-P2 and T complex morphology and the spectrotemporal features of the words presented. Analysis of the complexes obtained showed that both complexes reflected the spectrotemporal features of the syllables. These results demonstrated the viability of these complexes for analysis of deficiencies in the cortical processing of speech, showing that they remained robust to natural variations in speaker and were consistent at the group level (and individual level for the P1-N1-P2 complex). Research into the effects of noise on these complexes may therefore have the potential to indicate levels of spectrotemporal deterioration at the cortical level. However, to the author's knowledge, no research has yet compared these complexes to behavioural measures of speech-in-noise performance.

2.2.2.5 Sentences and running speech

A recent development in evoked response measurement is the advancement in signal processing methods that allow for the analysis of responses to running speech. This advancement holds potential for the development of improved correlates to speech-in-noise performance as it allows for direct analysis of neural activity in response to natural speech. The use of stimulus with considerably greater ecological validity allows for the analysis of responses that better resemble the neural activity that occurs in everyday listening. Therefore correlates obtained from these responses may better represent deficiencies in the central auditory pathways that affect the processing of running speech than short and artificial stimuli. There are currently two similar methods that have been shown to be successful in the analysis of such stimuli: A brainstem-level correlation based approach which directly compares representations of the input stimulus to the response signal (Kong et al., 2015; Reichenbach et al., 2016; Forte et al., 2017) and a cortical-level regression modelling/prediction based method that estimates the transfer function of the auditory system given the known input (the stimulus) and the output (the EEG signal in response to stimulus) (Aiken and Picton, 2008; Lalor et al., 2009; Lalor and Foxe, 2010; Di Liberto et al., 2015; Ding and Simon, 2012). This section will provide an overview of these techniques and discuss the various studies that focus on or are relevant to the application of these methods for speech-in-noise estimation.

Another method, outlined in Kong et al., 2014, that may also yield useful correlates to speech-in-noise performance is the use of direct cross-correlation between the EEG signal and various stimulus derived features. By extracting features such as the envelope and f_0 from the speech stimulus and finding the time lag at which these features best correlate with the EEG, it is possible to determine if and at what stage in the central auditory pathways that the features are best represented. Work by Kong et al. (2014) used this method to study the cortical representation of the speech envelope for active and passive listening conditions. Participants ($N = 8$) were presented with active and passive listening tasks in noise and in quiet. Denoising source separation (DSS) was initially used to reduce the EEG signal to its most consistent components across multiple trials (150 tone pips were presented to the participant to obtain multiple trials for DSS). The most consistent component was then correlated to the envelopes of the attended and unattended speech signals. A variety of differences were found between the morphologies of the correlation functions for the attended speaker and the unattended speaker, suggesting a significant effect of attention on cortical representations of speech. This was built on by Power et al. (2011) to explore the relationship between spectral resolution of the attended speaker and cortical entrainment to that speech in the presence of a distracting speaker. Using an analogous method to the previous study, it was found that participant ($N = 8$, normal-hearing listeners) speech scores were significantly correlated ($r = 0.62, p < 0.001$) with the neural modulation index ("the RMS of the difference in the cross-correlation values between the attended and unattended speech over time lags" (Kong et al., 2015, p.788)). These studies demonstrate the potential for a correlation based correlate for speech-in-noise performance and further highlight the effects of attention on cortical tracking of speech. This method was further built on by Reichenbach et al. (2016) and Power et al. (2011). These studies focused on extracting the amplitude envelope modulated f_0 of the stimulus signal using a variety of methods. This signal is then correlated with single channel EEG to determine the correlation and time lag of the f_0 neural representation. Using normal-hearing listeners ($N = 14$) the study by Forte et al. (2017) demonstrated the presence of the fundamental frequency (and higher harmonics) in the brainstem response, consistently observing a correlation between response and stimulus at around 9 ms in all but 2 participants. It was also observed that this representation was modulated by the attention of the listener in the presence of competing speakers, with larger responses at latency peaks found for the attended speaker over the unattended speaker. This was true for 9 out of 14 subjects.

This research suggests that a correlation based method may be viable for producing speech-in-noise correlates, potentially with as few as 4 electrodes. It demonstrates the ability to extract acoustic features that are relevant to speech perception and shows that these features are encoded in the presence of competing noise. This research also highlights the necessary consideration for attention which is shown to be a factor even in subcortical speech processing. However, there are a number of discrepancies between this approach and other research that should also be considered: It is noted by Reichenbach et al. that reduction in f_0 representation was observed when speech is intelligible, in contrast to previous research by Galbraith et al. (2004) which observed the opposite effect. It is suggested that this may be as a result of discrepancies between responses to running speech and repeated stimulus. This therefore may contrast the findings of studies using short, repetitive stimuli discussed in Section 2.2.2.4, which suggest that f_0 is less likely to be useful as a speech-in-noise correlate, as recognition performance is thought to be modulated to a greater degree by transitions in speech than in steady state sections. Other considerations include the need to present repetitive stimuli prior to running speech in order to utilise DSS.

An alternative method that has been the focus of recent research is the use of a temporal response function (TRF, or mTRF for multivariate-TRFs) for analysis of running speech. This method can be thought of as an extension of the correlation method presented previously, that attempts to model the auditory system as a linear time-invariant system, using linear regression. This model assumes that the output of the system ($y(n)$, the EEG data) can be described as a convolution ($*$) of the input of the system ($x(n)$, the stimulus) with an impulse response ($w(n)$), plus residual noise (Lalor et al., 2009):

$$y(n) = x(n) * w(n) + \text{Noise}.$$

This can be thought of as a regression model, where a least-squares solution is used to estimate an impulse response, that minimises the error between input and output. This response can then be used for estimation of unseen EEG data from an input signal, or for reconstruction of an unseen input signal from EEG data. This method has been used in a number of studies, particularly for the purposes of understanding the level to which different representations of the input signal are present in cortical responses to running speech. Studies have shown that the amplitude envelope (Aiken and Picton, 2008; Lalor and Foxe, 2010), spectral representations and phonemic representations (Di Liberto et al., 2015; Di Liberto and Lalor, 2017) may all contribute to neural processing using this method of comparing a reconstructed signal to an unseen original.

Studies have also used the TRF method when focusing on understanding the localisation of different representations throughout the cortical central auditory pathways. Using this method Brodbeck et al. (2018) demonstrated the ability to localise the sources of representations from lower level acoustic processing (amplitude envelope) to mid-level (word frequency) and cortical level semantic processing (semantic composition). By generating source estimates, creating TRFs for each individual estimated dipole, and recombining these TRFs, a model was created that described the activity in response to stimulus, both temporally and spatially. Results supported previous understanding of the hierarchical composition of the central auditory pathways, in addition offering further insight into the possible sources of speech processing at various levels. Additionally, these models were generated with as little as 6 minutes of MEG data, suggesting that this method may yield benefits for reduction of stimulus needed for generating TRFs. It is noted that the localisation techniques used are limited by the relatively low number of sensors. This suggests that this method may be less applicable to EEG measurements, which typically use fewer sensors than MEG.

There are limitations to the current implementation of the mTRF that should be acknowledged when considering its use for running speech analysis. A commonly highlighted issue is its assumption that the modelled system is linear and time-invariant, despite it being well understood that the auditory system is neither (Lalor et al., 2009; Power et al., 2011; Brodbeck et al., 2018; Vanthornhout et al., 2018). A potential solution to this issue is to substitute a non-linear model such as an artificial neural network or source vector machine, allowing for non-linearities in the system to be better represented in the model. This was explored by Power et al. (2011) who found only minor improvements through a non-linear extension based on a Volterra series. Given that this model only used a second-order extension, there is still potential for further improvements to the model. A further consideration is the relatively poor correlations that are achieved between the original stimulus and the reconstruction produced by the model (maximum correlations can be as little as 0.2 in the presence of competing speaker (Ding and Simon, 2012)) as well as the large inter-subject variability in results. These aspects will likely affect the robustness of these models when considering individual listener's responses and may result in the need for long data collection sessions to generate reliable results.

It is clear that the TRF is a promising method for assessing speech-in-noise performance as it allows for direct analysis of speech feature representation along the auditory pathway. Vanthornhout et al. (2018) investigated the potential for use of the TRF in this context. This study focused on correlating reconstructions of the speech envelope in response to running speech with results from a Matrix test. Normal-hearing listeners ($N = 24$) were presented with concatenated matrix test stimuli in up to 7 SNRs and 15 minutes of running speech. EEG data was recorded in response to these stimuli and an TRF model was trained using the running speech stimulus only. Envelope reconstruction was performed using the concatenated Matrix test responses, with correlation to the original clean matrix speech used to compare with the behavioural measure. Results were promising, showing that the cortical representation of the speech envelope could be better reconstructed as the SNR increased. This was also found to correlate well ($r = 0.69, p = 0.001$) with SRTs obtained from the behavioural test. This research was further developed by Lesenfants (2019), extending the method to use higher and lower level phonetic and spectral features proposed by Di Liberto et al. (2015). The model used for prediction is also adapted to use a grand-average TRF model, rather than individual models for each participant. Results suggested that using a combined spectrogram-phonetic representation in the theta band could provide SRT estimations within 2 dB SNR for more than 80% of the normal-hearing participants. The authors of this study conclude that this predictive performance provides an objective measure of speech-intelligibility, but that further research is needed to assess the measures performance over an extended cohort of participants.

A number of methodological aspects should be considered that may be addressed to build on this work: The studies used a small number of normal-hearing listeners, with only minor variations in performance on the behavioural test. Participant's SRTs were in the range of -10 dB SNR to 0 dB SNR. Given the small range of subject performance levels and the matrix test's reported inter-subject variability of 0.5 dB (Luts et al., 2014), it is not clear if this method would generalise to larger populations and to those with varying degrees of speech-in-noise performance. As noted for previous evoked response methods, results are presented at the group level, so further research would be needed to understand the individual predictive performance of these methods. Additionally, as these methods focus on cortical processing, cognitive processing may influence the results. Use of alternative acoustic features such as the temporal fine structure could be used to give further insight into processing that occurs at the brainstem (Maddox and Lee, 2018). This would likely also be less influenced by such higher level cognitive function. As suggested previously, the use of higher level models may improve overall reconstruction accuracy. Finally, the model used for prediction could be extended to a non-linear model to improve overall correlation between stimulus and reconstruction.

2.3 Traditional interventions to improve speech-in-noise performance

When surgical and/or pharmaceutical interventions do not provide an effective solution for a person's hearing loss, external neuroprosthetics are typically used in order to aid in speech-in-noise performance. There are a wide variety of clinical interventions available, which are selected based on the type and severity of the hearing loss. There are a number of challenges in the design of these devices, both acoustically (devices are required to process a wide range of input sound levels and frequencies), and practically (devices must provide a robust intervention for day-to-day use in a variety of environments) This sections will provide a brief overview of available hearing assistive technologies, focusing on typical clinical interventions such as acoustic or bone-anchored hearing aids (Section 2.3.1), and cochlear implants (Section 2.3.2).

2.3.1 Hearing aids

For many hearing losses, the most appropriate intervention is currently an acoustic hearing aid, which amplifies sound to compensate for loss of hearing sensitivity. Recent improvements in design and manufacturing techniques have lead to a variety of hearing aid styles, as discussed in Section 2.3.1.1. Hearing aid designs must account for many inherent restrictions, including limited power supply, the need to be very small, and the variety of noises and distortions that degrade a listener's speech-in-noise performance. The advent of modern digital technology in recent years has allowed for the development of increasingly complex signal processing strategies. These aim to improve speech-in-noise performance by enhancing the quality of the acoustic signal received by the hearing aid, as discussed in Section 2.3.1.2. These sections focus on the impact of hearing aid design decisions on speech-in-noise performance outcomes for hearing impaired listeners. Due to the wide range of designs and implementations, this section will provide an overview of the most common concepts, with references to detailed reviews provided.

2.3.1.1 Hardware

Hearing aids are designed with a broad focus on improving the quality of user outcomes (including improvement to speech-in-noise performance), as well as the durability of the device and its aesthetics (Moore and Popelka, 2016). In modern hearing aids a complex combination of hardware is used to maximise performance in these areas. These devices typically consist of the following components:

- Microphone(s) used to transduce speech and environmental sounds to electrical signals.
- An analog-to-digital converter, used to convert analog electrical sound signals to a digital representation for further processing and noise reduction.
- Sets of application-specific integrated circuits (ASICs) and digital signal processing chips, used to perform noise reduction and other signal enhancement techniques (discussed in Section 2.3.1.2).
- Increasingly, hearing aids may include a wireless module, allowing for communication with other devices and/or hearing aids via radio/bluetooth.
- A receiver (either with or without an accompanying digital to analog converter), used to convert the processed audio signal back to audio for presentation to the

listener. Depending on the hearing aid type, this may be a small speaker or bone-conduction shaker.

- A small ergonomic case designed to house the device components discreetly.
- A battery, used for powering the electronic components of the hearing-aid.

This section will provide a brief outline of the key hardware manufacturing advances that have had a significant impact on speech-in-noise performance. For a comprehensive overview, please refer to Fay et al. (2016).

Manufacturing has progressed in almost all aspects of hearing aid hardware over the past 30 years. In recent years, the focus of hearing aid design has been primarily on improvement of speech recognition performance, on improved overall sound quality, reduced power consumption (for longer battery life) and on reduction in size (Hänsler and Schmidt, 2006; Launer et al., 2016; Killion et al., 2016; Moore and Popelka, 2016). Development of a discrete solution is necessary due to the perceived stigma of using a hearing aid (Bartkiw, 1988; Parette and Scherer, 2004; Erler and Garstecki, 2002), but presents a trade-off between device size and potential audio processing performance. Advancements in transducer, signal-processing and wireless technologies have enabled the development of a wide range of different hearing aid styles. Modern digital hearing aids may be mounted visibly in the ear, completely in the ear (not visible), behind the ear or mounted externally on the mastoid in the case of bone conduction hearing aids. However speech-in-noise performance remains an issue for all of these device types.

A key area of improvement has been in transducer technologies. The development of higher quality transducers has improved bandwidth and frequency responses, as well as improved resistance to shock damage, magnetic shielding, and microphone insensitivity to vibration (Killion et al., 2016). This has led to reduced additive mechanical and electrical noise, as well as less distortion of the speech signal. Modern transducers are now able to capture and reproduce speech signals with imperceivable levels of distortion and broadband flat frequency responses. Any further gains in this technology would be largely eclipsed by the quality of the ear-mold fit, and the limitations of battery power on the audio amplifiers of the hearing aid (Lewis and Moss, 2013; Killion et al., 2016).

The advent of digital signal processing and ASICs have allowed for vastly more complex signal processing than was previously possible using analog circuitry alone. These chips provide increased power efficiency and can be flexibly reprogrammed. This allows for adaption of algorithms for different environments and fine-grained tuning based on user's hearing loss (Hänsler and Schmidt, 2006). However, given the small form factors required for hearing aids, signal-processing strategies are still fundamentally limited by the need to perform processing in real-time on relatively low-powered chips. There are a plethora of algorithms implemented on these chips that aim to improve speech-in-noise performance - The most prominent algorithmic developments are detailed in the following Section 2.3.1.2.

Development of wireless technologies such as bluetooth and low-frequency radio transmission have allowed for innovation in hearing aid design. Wireless modules are commonplace in modern hearing aids, allowing users to connect to media devices such as computers and mobile phones. Direct connection to these devices has been shown to be particularly advantageous for increasing the SNR to improve speech recognition performance (Kim et al., 2014; Chen et al., 2021). Additionally, bilateral hearing aids may now communicate with each other, allowing for the implementation of more advanced binaural signal-processing strategies (reviewed in Section 2.3.1.2).

Although these technological advances are promising, there are still many limitations that restrict improvements to speech-in-noise performance. These currently include the need for ever smaller device form factors, limited battery power which restricts signal processing capabilities, and the required simplicity of wireless signals between devices (due to the limits of small wireless modules).

2.3.1.2 Signal-processing strategies

The DSP and ASIC chips (described in above Section 2.3.1.1) can be programmed with a plethora of signal processing techniques, designed to cleanup and amplify audio. Hearing aid signal-processing algorithms typically fit into one of the following categories (Launer et al., 2016):

- Noise reduction
- Environment classification
- Frequency- and level-dependent amplification for restoring audibility and acceptable loudness.

Each stage of digital signal processing incurs a delay to the signal. A delay of more than 10-12 ms causes an unwanted echo, therefore the complete signal processing strategy of a hearing aid must total less than this (Stone et al., 2008). This is a primary limiting factor on the performance of hearing aid signal processing techniques. This section provides a brief overview of the most common signal processing strategies and their effect of speech-in-noise performance. For a comprehensive overview, please refer to Launer et al. (2016).

Hearing aid digital signal processing strategies start with a frequency analysis (either FFT/IFFT or filterbank based). This is key stage that supports much of the signal processing techniques that follow. There are various trade-offs between frequency analysis methods but all are required to work within small windows to reduce latency. This analysis is therefore usually very limited in terms of frequency resolution (for example, for a 16 KHz samplerate a typical FFT size of 128 samples results in frequency bins spaced coarsely at 125 Hz intervals), limiting the quality of subsequent algorithms (Launer et al., 2016; Hänsler and Schmidt, 2006).

Multiple techniques are employed to reduce the unwanted background noise in a signal. Multi-mic and binaural beamforming algorithms are commonly used to direct the focus of microphones towards the target speaker. Many other algorithms are also used to remove stationary noises, impulsive sounds and to reduce feedback from the receiver. Detailed description of these algorithms are outside the scope of this review — for more information refer to Launer et al. (2016). The effectiveness of noise reduction algorithms on speech-in-noise performance varies, with considerable benefit shown in particular for beamforming based approaches. Picou et al. (2014) shows reduced listening effort and improved speech-intelligibility, particularly for binaural beamforming approaches. However beamforming techniques are only effective when the source and noise are in separated spatially. The benefits of other noise reduction techniques are less clear. Although algorithms show promise for specific environments, many don't provide significant benefit, particularly when the competing noise is another speaker (Ricketts and Hornsby, 2005; Bentler et al., 2008)

As hearing aids are required to operate in a wide variety of environments, their signal-processing strategies must be able to adapt accordingly. To do this, a higher-level heuristic understanding of the environment can be used to inform signal processing strategies (For example, an ideal signal-processing strategy would be very different for listening in a quiet environment than for listening to speech in a noisy restaurant). This is achieved using some form of environmental classifier. These systems work by analysing the statistical properties of the incoming sound, which then typically use machine learning based classifiers in order to make decisions on the type of environment (Nordqvist and Leijon, 2004). Little direct evidence suggests that these algorithms improve speech-in-noise performance. However, the ability to accurately predict environment allows for more specialised signal processing strategies, which is likely to have a positive impact for speech-in-noise performance (Bentler and Chiou, 2006).

Frequency and level dependent amplification is commonplace in all digital hearing aids. This uses the frequency analysis described above with per-channel dynamic range compressors. These algorithms compress, limit and expand the signal on a per-band basis to control the perceived loudness of a signal, aiming to counter loudness recruitment by maintaining an optimal listening level for speech and sound recognition. Gain and compression settings are set by an audiologist based on the user's PTA scores, but may also take the user's preferences in to account. Gains may then be adapted during use based on an environment classification algorithm, as described above. The effect of dynamic range compression on speech-in-noise performance strongly depends on the settings of the specific algorithm and is highly user dependent, requiring careful tuning to avoid distortion of the speech amplitude envelope. An in depth review of compression strategies is provided by Launer et al. (2016).

In addition to level compression, frequency compression/transposition is sometimes used to restore high frequency hearing. These algorithms remap higher frequency components, either by superimposing them onto lower ranges or by compressing the entire spectrum of sound. Various methods have been proposed, as reviewed in Simpson (2009), Alexander (2013), and Picou et al. (2015). Overall performance benefits using frequency compression are modest, with severely hearing impaired participants gaining the largest benefits.

2.3.2 Cochlear implants

A cochlear implant may be the most appropriate intervention for listeners who suffer from severe to profound hearing losses, and so would not benefit from an acoustic hearing aid. A cochlear implant uses many similar technologies to a hearing aid, with the exception of the receiver. A cochlear implant is a surgically implanted neuroprosthetic that bypasses the outer ear, stimulating the auditory nerve directly using and array of electrodes. This section will outline the hardware (Section 2.3.2.1) and signal processing (Section 2.3.2.2) components of cochlear implants and their impact on user outcomes for speech-in-noise performance.

2.3.2.1 Hardware

A cochlear implant has both an external and an internal component. The external component consists of microphones and a speech processing unit (analogous to the roles to the transducers and signal-processing units of a hearing aid. Refer to Section 2.3.1.1 for details on these components). Control signals derived from the input audio are then transmitted to the internal component via a radio, mounted using a magnet on the skull. The internal system consists of a receiver, and electrode array surgically inserted into the cochlea.

The electrode array is the interface that stimulates the auditory nerve, providing tonotopically mapped electrical pulses, which are perceived by the user as sound. Electrode arrays typically range from 16 to 22 electrodes used to stimulate the cochlea. This number is limited by the spread of electrical current along the cochlea, severely limiting the resolution of a cochlear implant in comparison to a healthy cochlea (which contains thousands of hair cells along the cochlea). Despite this limitation, cochlear implants have been shown to restore hearing to some of the most severely impaired listeners, in some cases resulting speech in quiet scores comparable to those of normal hearing listeners (Friesen et al., 2001). However, scores remain considerably poorer for speech-in-noise conditions. Studies have reported cochlear implants users performing in the range of 15 dB worse than normal hearing listeners in the presence of modulated noise. Additionally, percent words correct scores can also drop below 50% at even modest noise levels (Fu and Nogaki, 2005; Nelson et al., 2003; Fu et al., 1998a) due to the fundamental limitation of the maximum number of usable electrodes. However, techniques are currently being developed to try and limit the spread of current along the cochlea. The benefits of these for speech-in-noise performance are briefly discussed in the following Section 2.3.2.2, or see Carlyon and Goehring (2021) for an in-depth review. It should also be noted that the quality of the surgical procedure can have a substantial impact on speech-in-noise scores, however this is outside the remit of this review. For a review of modern surgical procedures and their impact on speech-in-noise performance refer to Lenarz (2018).

Additional benefits to speech-in-noise performance have been found in users who have received two implants. This provides access to additional localisation cues and binaural signal processing strategies (discussed in Section 2.3.2.2) that are not available to unilateral cochlear implant users. Although bilateral implantation is a growing trend, the majority of cochlear implant users are currently implanted unilaterally and cannot access these additional benefits (Dunn et al., 2008; Dunn et al., 2010; Laszig et al., 2004; Smulders et al., 2016).

2.3.2.2 Signal-processing strategies

Cochlear implant signal processing has benefited greatly from parallel research into hearing-aid technology. Algorithms such as automatic gain control, noise reduction and directional microphone steering follow similar strategies to those described in Section 2.3.1.2. An additional mapping strategy is also used to map audio down to the implant's array of electrodes. This section will review the most prominent signal processing strategies, and there benefits for cochlear implant user speech-in-noise performance.

As with hearing aids, noise reduction is commonplace in cochlear implants. However, given the generally poorer performance of cochlear implant users, even more traditional strategies have been shown to give user significant benefits. A range of traditional noise reduction algorithms have been investigated, with studies finding significant improvements of up to 2dB SRT or 25 percentage word recognition performance (Loizou et al., 2005; Hu et al., 2007; Dawson et al., 2011; Mauger et al., 2012; Ye et al., 2013; Chen et al., 2015; Wang and Hansen, 2018).

Directional microphone signal processing has also shown significant benefit, both in bilateral and unilateral formats. With approaches showing substantial improvements from 3.6 up to 16 dB SRT (Chung et al., 2006; Chung and Zeng, 2009; Spriet et al., 2007). As with hearing-aids these methods are only effective when the target speaker and noise are not co-located, and are also susceptible to degradation from reverberant environments.

There are 3 main strategies used for mapping audio to electrical stimulation: CIS, SPEAK and ACE. ACE has been shown to perform better in specific speech-in-noise tests (Kiefer et al., 2001) — Skinner et al. (2002) reported that for the CUNY sentence in quiet test, ACE performed on average 8.8% words correct higher than SPEAK and 5.6% words correct higher than CIS tests. However, ACE has also been shown to perform worse in consonant-vowel-consonant tests (performing 7–8% lower). Therefore overall it cannot be concluded to be a better strategy is the most effective for speech-in-noise performance. There have been both commercial and research based modifications of each of these strategies. Some aim to provide extra information such as the temporal fine structure (Wouters et al., 2015; Hochmair et al., 2015). Evaluations of these methods have not shown a consistent benefit for speech recognition or speech-in-noise performance (Magnusson, 2011; Riss et al., 2016; Riss et al., 2011; Müller et al., 2012). There are also many other experimental strategies — some provide additional biologically inspired processing before the strategy in order to mimic biological non-linearities, such as the Medial Olivocochlear Reflex (Lopez-Poveda et al., 2020). Others aim to enhance specific features of the input signal such as pitch modulations (Francart et al., 2015; Vandali et al., 2019).

Overall, developments in mapping strategies provide small but inconsistent improvements that are typically less than 10% in terms of improved percent scores, and around 1–2 dB improvements for SRTs. Noise reduction and other pre-processing methods have shown larger improvements, having been demonstrated to provide 2 dB reductions in SRT and improvements of up to 25% correct.

2.3.3 User outcomes

Hearing aids and cochlear implants have been demonstrated to be extremely successful interventions for improving patience' access to sound. Cochlear implant are widely recognised as one of the most effective neuroprosthesis developed to date — allowing the severely hearing impaired to achieve performance in excess of 80% words correctly identified for high-context sentences ("NIH Consensus Conference. Cochlear Implants in Adults and Children" 1995). Given the challenging listening environments faced by a hearing aid, they too provide considerable improvement, particularly in quiet listening conditions (Souza, 2016). This is a remarkable improvement that is as a result of development in these technologies, as discussed in previous Sections 2.3.2 and 2.3.1. However, for speech-in-noise performance, both hearing aids and cochlear implants provide significant improvements for hearing impaired listeners (Souza, 2016; Carlyon and Goehring, 2021). Neither hearing aid or cochlear implant users are able to perform as well as normal hearing listeners, with cochlear implant users experiencing the greatest difficulties. Kaandorp et al. (2015) assesses the 3 groups using sentences and digit-triplets in noise. For sentences in noise the normal hearing listeners scored -4.2 dB SRT on average (std=0.8). Hearing aid users scored significantly worse at 2.1 dB SRT (std=4.8) and cochlear implant users score worst at 8.0 dB SRT (std=6.1). For digits in noise, the normal hearing group scored -9.3 dB SRT on average (std=0.7). Again hearing aid users scored significantly worse at -4.4 dB SRT (std=3.5) and cochlear implant users score the worst at -1.8 dB SRT (std=2.7). These results highlight the limits of current interventions to provide adequate speech-in-noise performance for hearing impaired listeners. These limits are likely a combined result of the fundamental technological limits of the interventions (such as the limit resolution of cochlear implant electrode arrays) and of the limits of the fitting procedures (such as the use of predominantly PTA scores for fitting of hearing aids).

2.4 Haptics to improve speech-in-noise performance

As discussed in Section 2.3, current interventions for addressing hearing-loss are limited in their ability to improve speech-in-noise performance. Recent research has demonstrated the efficacy of using haptic stimulation devices for the improvement of speech recognition (Summers, 1992) and speech-in-noise performance (Fletcher, 2021a). By providing missing sound information that cannot be perceived via the impaired auditory system via haptic stimulation, it may be possible to develop a new intervention that supports current clinical interventions. Section 2.4.1 will first evaluate the anatomical and physiological limits of the tactile system, which should be considered when designing such an intervention. Section 2.4.2 will then review previously developed tactile devices, with a focus on the hardware and software limitations as well as both experimental and clinical outcomes for users.

2.4.1 Anatomy and psychophysical limits of the tactile system

This section will first provide a brief overview of the anatomical mechanisms of the tactile system (such as the types of receptors found in the skin and the integration of tactile signals with auditory signals in the brain) and the psychophysical properties of tactile perception to be considered when designing a haptic intervention. This section will then outline the sensitivity of the system with regards to intensity, temporal response, and frequency changes as well as spatial acuity. Finally this section will discuss the perception of complex stimuli, with comparison to the auditory system.

There are two types of skin: glabrous and hairy skin. Glabrous can be found on areas such as the palm, fingertips and lips, and hairy across most other parts of the body.

Each contain various densitys of the four primary mechano-recoptors: Merkel's disks, Ruffini endings, Meissner's corpuscles and Pacinian corpuscles. Merkel's disks, Ruffini endings and Pacinian corpuscles are found in both hairy and glabrous skin. In addition to these, Meissner's corpuscles are also found in glabrous skin. These receptors work in conjunction to form the overall perception of tactile stimuli. However, the sensitivity to stimulus varies based on the types and density of receptors at the location on the body. This variability should be considered both when interpreting the following presented limits of the tactile system, and when selecting a suitable body location for the intervention. The following will focus on the perceptual limits that have been well established for key areas such as the hand and forearm. For an in depth text on this area, refer to Gescheider et al. (2010).

Tactile responses from receptors travel via the somatosensory pathways to various nodes throughout the brain. Extensive connections have been shown between tactile and auditory pathways, from lower areas beginning at the cochlear nucleus, up to cortical regions (Aitkin et al., 1981; Foxe et al., 2000; Shore et al., 2003; Shore et al., 2000) Physiological studies have also demonstrated interaction between the pathways, showing that activity in the auditory cortex can be modulated via haptic stimulation (Lakatos et al., 2007; Meredith and Allman, 2015). The existence of these connections suggests that there is considerable potential for audio-haptic integration. This is further supported by imaging studies (Kassuba et al., 2013; Schürmann et al., 2006) and psychophysical evidence of audio-haptic integration (reviewed by Fletcher (2021a). However, the psychophysical limits should be considered to design an intervention that maximally utilises these mechanisms.

Of the perceivable features of tactile stimuli, the tactile system is particularly sensitive to changes in intensity. For the hand and index finger, differences of as little as 1.5 dB between successive stimuli can be detected (Craig, 1972; Gescheider et al., 1996), and similar results have also been shown for the wrist (Fletcher et al., 2021a; Summers et al., 2005). This is similar to that of the healthy auditory system (Harris, 1963; Penner et al., 1974; Florentine et al., 1987), suggesting that across ear intensity difference cues may be well substituted via tactile stimulation for the hearing impaired (Fletcher et al., 2020a). The system's dynamic range and resolution also show promise for transfer of speech cues. The tactile system has a dynamic range of 55 dB (Verrillo et al., 1969), which although poorer than the healthy auditory system (which typically has a range of 115-130 dB), is particularly sensitive in comparison to a cochlear implant, which can only provide between 10 and 20 dB of range (Zeng and Galvin, 1999; Zeng et al., 2002). The resolution within this range also provides up to 40 steps of individually discriminable intensities, far greater than the maximum of 20 dB for a cochlear implant (Gescheider et al., 1996; Kreft et al., 2004; Galvin and Fu, 2009; Fletcher et al., 2021a; Fletcher et al., 2021b).

Unlike intensity, the tactile system's sensitivity to temporal changes is limited when compared to the auditory system. Gap detection thresholds for tactile stimulation are around 10 ms on average (Gescheider, 1966; Gescheider, 1967), consistently higher than the 2-5ms threshold reported for normal-hearing, hearing-impaired and cochlear implant users (Moore and Glasberg, 1988; Garadat and Pfingst, 2011; Plomp, 1964; Penner, 1977). In addition, competing maskers have been shown to raise stimulus detection thresholds when the masker precedes the target stimulus by as much several hundred milliseconds (Gescheider et al., 1989). This contrasts the auditory system that typically is not affected by a preceding masker of more than 100 ms (Elliott, 1962; Shannon, 1990). The poor gap detection and masking performance of the tactile system precludes the perception of much of the temporal fine-structure of speech directly. However, the relatively slowly modulating speech amplitude envelope has been delivered via haptic stimulation in a number of studies, providing considerable benefits — in particular for those with cochlear implants (Fletcher et al., 2018; Fletcher et al., 2019; Fletcher et al., 2020b; Proctor and Goldstein, 1983; Spens and Plant, 1984). The tactile system is well suited to providing this speech feature. This is due to its high sensitivity to envelope modulations at frequencies that are important for speech recognition (Weisenberger, 1986; Drullman et al., 1994). Given the particularly poor dynamic range and spectral resolution of a cochlear implant, these users are a set that would therefore be expected to see the greatest benefit from this type of haptic stimulation.

Both on the finger and on the forearm, the tactile system is considerably poorer at discriminating changed in frequency than the auditory system. It is only able to discriminate differences of 20% at 50 Hz and of 35% at 200 Hz (Goff, 1967; Rothenberg et al., 1977), compared to the healthy auditory system's ability to discriminate differences of down to 0.6% for frequencies between 0.25 and 2 kHz (Sek and Moore, 1995). The auditory system's performance is poorer for cochlear implant users, due to the spectral resolution caused by the implant. However, cochlear implant users are still able to detect frequency changes of 10–25% at 500 Hz and 10–20% at 4 kHz (Turgeon et al., 2015). The bandwidth of the tactile system is also reduced, ranging from 30–1000 Hz, compared to the 20– 20,000 Hz (Suzuki and Takeshima, 2004) range of the healthy auditory system. Vocoder based approaches have been used to remap frequency ranges to account for this (Fletcher et al., 2018; Fletcher, 2021a). These characteristics suggest that a direct mapping of auditory frequency components to haptics may be of limited suitability to provide fine-grain pitch cues.

An alternative approach for mapping frequency content is to map this feature spatially across locations on the skin. This approach is limited primarily by the design space (dependent on body location), ability to separate actuators within the space, and the spread of excitation between tactile transducers of the device. For the hand, localisation accuracy is thought to be high, with discrimination threshold measures reported at 0.5 mm to 6 mm (Johnson and Phillips, 1981; Perez et al., 2000). Higher thresholds are reported for the forearm, at between 25 mm to 50 mm (Cholewiak and Collins, 2003; Schatzle et al., n.d.). It should be noted that discrimination of multiple sites is highly dependent on stimulus type (Boldt et al., 2014; Perez et al., 2000) and non-standardised methodologies may result in inaccuracies in previously reported spatial resolution (Johnson and Phillips, 1981; Craig and Johnson, 2000). It should also be considered that the perceptual interaction of stimulation at multiple locations on the skin is complex and only partially understood - effects such as masking, summation, suppression and enhancement have all been observed when stimulation occurs at multiple sites and should be accounted for when designing a new haptic intervention (Fletcher and Verschuur, 2021) — See Summers, 1992, p.19-23 for further details. Overall, the data presented suggests that actuators could be effectively arranged across the skin to provide a spatial mapping for auditory features. Success has already been found using spatial mappings for sound localisation (Fletcher et al., 2020a; Fletcher et al., 2021b; Fletcher et al., 2021a; Fletcher et al., 2020b), and in commercial haptic aids (see the following Section 2.4.2 for further details). The spatial mapping of pitch and amplitude envelope features will be discussed further in Chapters 5 and 6.

2.4.2 Tactile stimulation devices

Over the past century many devices have been developed, aiming to provide tactile stimulation for the hearing impaired. This section will provide a brief overview of the most prominent commercial and research based devices. The practical limitations, both in hardware (Section 2.4.2.1) and signal-processing strategies (Section 2.4.2.2) will be outlined and recent advances in technology that may support development will be highlighted. For further information on this topic please refer to the latest reviews by Fletcher (2021a), Fletcher and Verschuur (2021), and Fletcher (2021b) and Summers (1992)

2.4.2.1 Hardware

Much of the technology needed in the development of a haptic device follows that of a hearing aid or cochlear implant. Components such as microphones, analog-to-digital converters, wireless modules and batteries all use similar technology. This section will review any notable differences to that of hearing aids, but for an overview of these technologies please refer to Section 2.3.1.1. The additional components that must be considered in the design of a haptic device are:

- Motors - to provide haptic stimulation on the skin
- Motor drivers - to calibrate and provide driving currents to the motors based on control voltages of the signal processing strategy
- Component casing and mounting on the skin

The previous Section 2.4.1 focused on properties of the finger, hand, wrist, or forearm (due to limited data for other locations). However, tactile aids have been designed for use at many locations on the body. These additionally include the abdomen, sternum, and back. Table 2.4 provides a list of some notable tactile aids for the hearing impaired, alongside information on their physical location. Locations such as the hand are popular in lab research due to the high sensitivity of the skin at this area (Huang et al., 2017; Huang et al., 2019; Fletcher et al., 2018). However, commercially viable devices typicaly opt for areas such as the wrist or back due to the practical restrictions when performing daily tasks with devices mounted to the hand (Summers, 1992, p.9-11). Additionally, the size and appearance of the device should be cosidered. For hearing-assistive devices the stigma of such a device is a significant factor inhibiting their uptake (Bartkiw, 1988; Bispo and Branco, 2008; Erler and Garstecki, 2002; Parette and Scherer, 2004). It is possible that the same will be true of tactile devices (Fletcher, 2021b).

Device name	Body location	Shaker type	Audio input type	Audio sensitivity	Source
Siemens Minifonator	Wrist	1 "Electro-mechanical transducer"	External mic	100 Hz–3 kHz	Weisenberger (1989)
Tactaid II+	Sternum or top/bottom of wrist	2 V1420 Variable-reluctance tactile motors (origin unknown)	External mic, built-in mic or T-coil	100 Hz–8 kHz	Tactaid II+ Fitting Manual M-3 (1992)
Tactaid VII	Wrist or abdomen	7 motors (type unknown)	External mic, built-in mic or T-coil	200 Hz–7 kHz	Galvin et al. (1999)
Neosensory Buzz	Wrist	5 LRA motors	Built-in microphone	300 Hz–3 kHz	Neosensory (2020)
TASBI	Wrist	5 LRA motors	None (external conversion to tactile signal)	300 Hz–3 kHz	Pezent et al. (2019)
Vibrotactile vest	Chest, back and abdomen	ERM (model #307-100 from Precision Microdrives)	Arduino Uno/External conversion	Unknown	Novich and Eagleman (2015)
Soft-talker	Wrist	DC driven "Micromotor" - Portescap (UK) 712	Built-in microphone	150 Hz - 10 kHz	Walker et al. (1987)
Queen's University Tactile Vocoder	Forearm	18 solenoids	None (external conversion to tactile signal)	160 Hz - 8 kHz	Brooks et al. (1985)

Table 2.4: Summary of some notable haptic devices

As with hearing-aids and cochlear implants, microphones placement is key for providing the clearest representation of speech. Current and historic haptic devices do not use behind the ear devices, typically opting for onboard or external microphones (see Table 2.4). This simplifies the design processes, as it does not require communication between behind the ear devices and the haptic device. However this can result in issues such as wind noise and noise from friction between the microphone and user's clothing (Fletcher, 2021a; Fletcher and Verschuur, 2021). Recent work by Fletcher et al. (2020a) has also shown improved sound localisation performance when presenting binaural audio via haptic stimulation. These binaural cues, recorded from behind the ear microphones, would not be available to provide localisation cues when using built-in microphones. Use of behind the ear devices is an area that may provide considerable benefit to haptic devices, and may soon be possible with the advent of modern wireless communication technology (discussed in previous Section 2.3.1.1).

In place of the receiver or electrode array for traditional hearing interventions is an array of motors or electrodes used to stimulate the skin. The choice of transducer is crucial, with properties such as frequency and amplitude response, power consumption, and form factor needing consideration. The three most prominent transducer types are Eccentric Rotating Mass motors (ERMs), Linear Resonant Actuators (LRAs) and Piezos. ERMs provide high intensity vibration for stimulating the skin (approximately 1-7 G for representative motors of an appropriate size (Precision Microdrives, 2021a; Precision Microdrives, 2021b; Precision Microdrives, 2021c)). However, these motors have coupled amplitude and frequency so an increase in one results in an increase in the other. As mentioned in previous Section 2.4.1, formal assessment is needed to better understand the effects of using complex stimuli such as this. LRAs provide less force than ERMs (Approximately 0.7-1.5 G for representative motors of an appropriate size (Precision Microdrives, 2021e; Precision Microdrives, 2021d; Precision Microdrives, 2021f)) but have decoupled frequency and amplitudes, vibrating at a single resonant frequency. Piezos are able to provide decoupled frequency and amplitude, as well as high output force (Approximately 2.5—36 G for representative motors of an appropriate size (TDK, 2021d; TDK, 2021f; TDK, 2021a; TDK, 2021b; TDK, 2021c; TDK, 2021e; TDK, 2021h; TDK, 2021g)) but are considerably more expensive than the alternatives. For any motor type the contactor size should also be considered, as increased contact with the skin is known to increase the sensitivity due to stimulation of increased numbers of receptors (Verrillo, 1963).

In addition to motor selection, control voltages from the signal-processing strategy must be effectively converted to driving currents for the given motors. These drivers must account for the pressure placed on the motor, as well as the motors start-up and stopping times in order to faithfully reproduce the stimulation from the signal-processing. There are many driver technologies that have been developed with advanced techniques such as automatic calibration, active braking (the technique of applying an inverse voltage to a motor to decrease motor intensity more rapidly) and overdrive (the technique of applying a voltage higher than the target intensity to a motor to increase motor intensity more rapidly) to optimise the motor's performance (Texas Instruments, 2021; Boreas Technologies, 2021).

Finally, the form factor of the device should be designed appropriately such that it will not deter the user. Ideally devices should be discreet, comfortable and not interfere with the user's day-to-day activities (Fletcher and Verschuur, 2021; Summers, 1992). This is an area that can be improved on over traditional devices such as the tactaid and soft-talker tactile aids illustrated in Figure 2.4. The reduction in size of signal-processing units, batteries and haptic drivers should allow for more practical devices than was possible during the design of devices such as these.

There are a large number of research and commercial devices that have been developed to explore the benefits of haptic stimulation. Some notable devices have been listed in Table 2.4. There are also many examples of custom made/adapted devices that have been used for research purposes, such as the adapted HV Labs shakers used by Fletcher et al. (2020a), Fletcher et al. (2019), and Fletcher et al. (2018) and the finger-mounted tactaids used by Huang et al. (2017) and Huang et al. (2019). These devices would need considerable modification for real-world application.

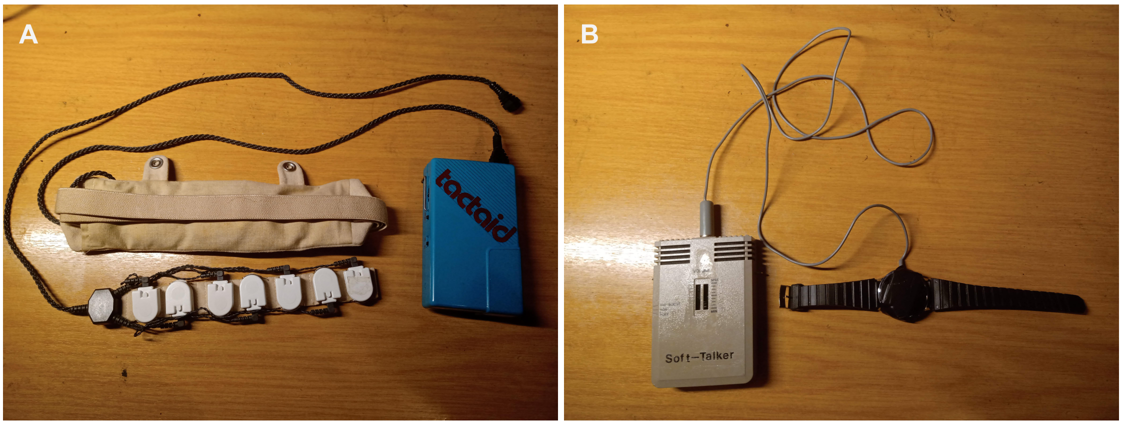

Figure 2.4: Image of the Tactaid VII haptic aid (A) and the Soft-Talker haptic aid (B)

2.4.2.2 Signal-processing strategies

Unlike hearing aids and cochlear implants, there is no consensus on the most appropriate strategy for mapping audio to haptics. A number of strategies have been proposed, that can be broadly categorised into two categories: vocoder based strategies and feature based strategies. Vocoder based strategies typically split the input audio into frequency bands, extract the envelope, and then use these envelopes to modulate a motor or set of motors. Examples of this approach include the Tactaids II (Tactaid II+ Fitting Manual M-3 1992) and studies by Fletcher et al. such as Fletcher et al. (2018) and Fletcher et al. (2020b). These strategies aim to provide a sense of the broad spectral shape of the speech signal over time. However, the need for wide bandwidth filters across the frequency range of interest results in the loss of key speech features such as the fundamental frequency. Despite this limitation, promising results have been shown in particular for cochlear implant users. Despite these limitations Brooks et al. (1985) reports that, with extensive training, in quiet conditions a vocabulary of up to 250 word can be learnt using a vocoder based method. Fletcher et al. (2019) shows improvements in noise, reporting improvements of 8.3% performance points for correctly identified words in noise, with some users improving by as much as 20%. Further improvements are also demonstrated when the noise is separated from the speech in space; Fletcher et al. (2020a) reports improvements of 2.8 dB improved SRT when the noise was ipsilateral and 2.6 dB improvements when the noise was presented contralaterally.

Feature based methods aim to address this by explicitly extracting higher level representations of key speech features such as the fundamental frequency (Huang et al., 2017; Huang et al., 2019), speech formants (Blamey and Clark, 1985) and voicing of the speech (Ifukube and Yoshimoto, 1974; Bin Afif et al., 2019). Using a mapping of extracted fundamental frequency to tactile frequency Huang et al. (2017) reported improvements of 2.2 dB SRT on average, with some users improving up to 7 dB. However, this study bypassed the noise by presenting clean speech to the signal-processing strategy. Further work will be needed to determine the robustness of this strategy to noise. These studies demonstrate the efficacy of both signal-processing strategies to offer improvements to speech-in-noise performance, however further research is needed to better understand which features are best utilised by the user to gain the improvements reported.

Many studies haven't addressed the issue of background noise, choosing to provide the clean signal directly to the haptic signal processing. This is a non-trivial issue in the design of haptic signal-processing strategies as the robustness of these techniques to noise would be important in the resulting performance of users in a real-world context. Approaches such as those developed in Fletcher et al. (2019) use additional noise reduction techniques in order to improve the SNR prior to presentation. The lack of an effect of haptics in Fletcher et al. (2020b) for the co-located noise condition (when no noise reduction was used) in comparison to effects in similar conditions with noise reduction in Fletcher et al. (2018) and Fletcher et al. (2019) suggests that these types of techniques are effective for haptic signal noise reduction. Further work is needed to assess the improvements that might be offered using more advanced noise reduction techniques, such as those presented in Section 2.3.1.2

Finally, as discussed in Section 2.3.1.2, hearing aids use both higher level features and information such as environment classifications to inform processing of the raw audio, forming a hybrid of the two categories outlined above. However, this form of contextual awareness has not been demonstrated in the literature for haptic devices. Translation of technologies such as this approach, as well as advanced noise reduction and wireless streaming from behind the ear devices are areas that hold much promise for the improvement of current signal processing strategies for a new haptic intervention.

2.5 Thesis overview

The aim of the studies presented in the following chapters is to provide two new methods for the assessment and treatment of speech-in-noise performance. The thesis is divided into 6 chapters, including the introduction and current background chapter. An outline of the following chapters is provided below.

Chapters 3 and 4 report results from experiments focused on the implementation and optimisation of a new objective neuroimaging method. This method provides a measure of how ecologically valid running speech is represented in cortical regions of the brain. This chapter forms the foundations for a technique that could provide additional insight, both for the underlying mechanisms of speech-in-noise performance and for the integration of haptic stimulation for the improvement of speech-in-noise performance. The data and analysis in these chapters provide novel methods for the improvement of an established method for analysing evoked responses. These chapters first validate previous approaches, providing an objective comparison of previously used methods. In addition, the chapters propose novel adaptions of previously used methods, accompanied by data which provides an objective evaluation of their performance. The results present compelling initial evidence for their efficacy as candidate methods to use for speech-in-noise measures.

Chapter 3 first addresses the most appropriate statistical model for applying the neuroimaging method, both in normal-hearing and listeners with a sensorineural hearing loss. Results show no significant difference between the two most widely used model types (SVD and Cholesky solver based linear regressions) in the current literature, but shows significant improvements for amplitude envelope reconstruction for the newly proposed ElasticNet model for the wide and delta analysis bands.

Chapter 4 then proposes an alternative perceptually-motivated feature for use with the selected model. Results suggest that the proposed perceptual loudness model can be more accurately reconstructed than the amplitude envelope for the wide and delta band models. These results could provide the basis for an objective method that, in the future, may be used for clinical tasks such as speech-in-noise performance assessment. Additionally the method may have further applications in areas such as analysing the underlying mechanisms of hearing losses and in analysis of audio-haptic integration.

The above work was published, in part in the conference proceeding by Perry et al., 2018, and Perry et al., 2019. These publications provided the academic community with insight into the potential for the proposed evoked response analysis method to be used for speech-in-noise performance, presenting initial experiment designs and pilot data to demonstrate this.

Chapter 5 details the development of a forearm-worn haptic intervention for the improvement of pitch discrimination in cochlear-implant users. The design choices made in development of the device are outlined and a new signal processing strategy for extracting and mapping pitch information to haptic stimulation is presented. This approach addresses a number of the shortcoming of previous haptic interventions (such as noise robustness of the signal processing strategy and limited frequency response of the skin), as discussed in Section 2.4.2. The approach is then tested for normal hearing listeners using a cochlear implant simulation, in a two alternative forced choice experiment. Tone complexes are used to determine the minimum pitch difference that can be discriminated with and without the haptic intervention. Results show that the users were able to discriminate pitch differences more accurately in the Haptic only and audio-haptic conditions than the audio alone condition. This section demonstrates the potential for a haptic intervention to provide key features of speech that would otherwise not be accessible for cochlear implant users.

This chapter was published in the peer-reviewed article Fletcher et al., 2020c. This article presented evidence supporting the use of a multi-channel haptic device to present pitch information with considerably greater resolution than is possible via a cochlear implant. This article is of significance to the academic community as it presents a methodology and accompanying data that could be used in the development of a real-world haptic intervention to improve speech-in-noise performance.

Chapter 6 builds on the results of Chapter 5, detailing the design of a compact wrist-worn intervention and signal processing strategy for the improvement of speech-in-noise performance. Participants are tasked with completing a standard speech-in-noise test, using either a new haptic processing strategy, an adapted strategy developed for previous haptic interventions or audio alone. Each condition was tested both before and after a substantial training regime. Results showed no significant effect of either haptic strategy on speech-in-noise performance scores. Discussion is provided on the unexpected lack of improvement found in this chapter and the areas that may require improvement are outlined.

Finally, Chapter 7 provides a summary of the reported findings, discussing their relevance to the current state-of-the-art in audiological and multi-sensory research. This chapter highlights the key questions that have resulted from the presented research and suggests future work that is needed. In particular, this section focuses on areas such as the potential for combination of the two methods presented to provide a tool for both basic research into the underlying mechanisms of audio-haptic integration and also as a clinical tool for an objectively optimised audio haptic intervention for the hearing impaired.

Chapter 3

Development of the Temporal Response Function model as a method for assessing speech-in-noise performance

Both clinically and in the lab, assessment of speech-in-noise performance is predominantly performed using behavioural speech-in-noise tests. However, these tests have limitations, such as the subjectivity of participant responses, test-time needed, and the lack of information that these tests provide on the underlying mechanisms of hearing loss. This places some limitations on the diagnostic information that can be extracted for clinical applications, or used to explore the underlying causes of hearing loss in the lab (An in depth review of the limitations of behavioural speech-in-noise tests is provided in Chapter 2.2.1. The TRF (described in Section 2.2.2) may have application in addressing the above issues. The TRF is a method that can be applied in response to continuous non-repeating speech. Recent advances have show that from the same collected data, analyses can be made at multiple levels of the auditory pathways, from brainstem (Maddox and Lee, 2018; Brodbeck and Simon, 2020) to cortex (Crosse et al., 2016; Di Liberto et al., 2018; Lalor and Foxe, 2010). Additionally this method can be performed without active participation from the patient, which may make it suitable for patients that are unable to respond directly (however, it should be noted that the method is sensitive to attention, a property of the method that is discussed further in Section 2.2.2). These properties make this method a suitable candidate for a new diagnostic and intervention fitting tool, that may also be particularly effective for groups such as children or those who cannot respond behaviourally. This method may therefore be able to compliment current classical speech-in-noise tests, and may offer greater insight into the neural underpinnings of speech-in-noise performance. However, there are currently a number of issues that should be addressed when considering this method for speech-in-noise performance analysis. One issue is that previous literature use a variety of different model implementations without formal assessment of the discrepancies between these. Assessment of model parity is therefore necessary to allow for comparison between previous studies, and to choose the most suitable implementation to achieve optimal predictive performance. Previously proposed implementations may also not be optimal in terms of reconstruction performance. Improvements made to these models could benefit factors such as test times and speech-in-noise performance prediction accuracy.

The following Sections will outline an experiment designed to validate previously proposed methods used for constructing the TRF and propose a new method that may provide improved performance. Much of the literature implements the TRF using one of two models: either a time-delaying ridge-regressor with a cholesky solver (Maddox and Lee, 2018; Gramfort, 2013) or with a singular value decomposition solver (Crosse et al., 2016; Riecke et al., 2019; Vanheusden et al., 2020) for computing the least-squares solution. The study aims to test the expectation that these implementations will provide similar reconstruction performance. Section 3.1.5.2 will outline the implementation of these models and will detail an additional regression model (the ElasticNet model) that may provide improved performance over previously proposed methods. These models will be assessed at 3 bandwidths — Wideband (0–20 Hz), Delta band (1–4 Hz) and Theta band (4–8 Hz). These bandwidths are commonly used for cortical TRF construction and are thought to be in the range of amplitude envelope frequencies that contribute most to speech intelligibility (Aiken and Picton, 2008).

Section 3.1 will outline a study which compares these measures to assess the performance of each for reconstructing amplitude envelopes. Results of the current study are then discussed in Sections 3.2 to 3.4, outlining the contributions made to the development of a method that could be applied both as an alternative speech in noise measure for diagnosis and as a measure to inform the fitting of interventions such as hearing aids, cochlear implants and haptic devices.

3.1 Methods

3.1.1 Participants.

This chapter is an analysis of existing data: the data used in the current study was originally collected by the authors of the study detailed in Vanheusden et al. (2020). The dataset contained evoked response and hearing assessment data for seventeen hearing-impaired participants (11 males, 6 females, aged between 64 and 70). Participants were native English-speaking and had a mild to moderate bilateral sensorineural hearing loss. Hearing profiles were obtained using PTA at audiometric frequencies from 250 Hz to 8 KHz. Hearing-loss profiles were as follows:

- 13 participants had bilateral sloping high frequency hearing losses

- 2 participants had flat bilateral hearing losses
- 1 participant had a bilateral ski-slope hearing loss
- 1 participant had a sloping hearing loss in their best ear and a flat hearing loss in their worst

Hearing thresholds for audiometric frequencies in each ear are detailed in the following Table 3.1

Frequency	Left ear	Right ear
250 Hz	23 ± 13dB	29 ± 21dB
500 Hz	23 ± 15dB	27 ± 21dB
1 KHz	29 ± 16dB	32 ± 23dB
2 KHz	43 ± 19dB	43 ± 21dB
3 KHz	54 ± 17dB	51 ± 20dB
4 KHz	61 ± 14dB	61 ± 17dB
6 KHz	69 ± 19dB	68 ± 24dB
8 KHz	68 ± 18dB	65 ± 18dB

Table 3.1: Average thresholds $\pm$ standard deviation for each ear at standard audiometric frequencies across participants

All participants were fitted with hearing-aids using the NAL-NL2 fitting procedure. The dataset contained data collected both when participants were aided by their hearing-aids and when unaided. For the purpose of the current study, only the unaided data was used. Responses to BKB materials in quiet were collected for further analysis. This data was also not used as part of this study. For further details, including the participant's hearing aid types and gain profiles, refer to Vanheusden et al. (2020).

3.1.2 Stimuli.

Twenty five minutes of running speech was used for collection of evoked response data. This was taken from a freely available audiobook (Colum, 2021), read by a female speaker. The audiobook was split into 8 segments of approximately 3 minutes each. The signals were then low-pass filtered at 3 KHz with a 120th order FIR filter. The stimulus was presented at 70 dBA equivalent sound pressure level (LeqA SPL).

3.1.3 Apparatus.

The stimulus was presented to the participant via a loudspeaker (placed 1.2 m in front of the participant) using a RME BabyFace soundcard (RME, n.d.) with a sample rate of

44.1 KHz. EEG data were collected using a BioSemi ActiveTwo EEG system (BioSemi, n.d.). 32 standard electrodes were used with two additional mastoid electrodes giving a total of 34 channels arranged following the 10–20 system layout. Evoked response data was recorded at a sample rate of 2048 Hz. Triggers were used for synchronisation of the stimulus output and evoked-response recording. These were produced via the soundcard, using an Arduino Due (Arduino, 2021) based trigger system with custom software to convert the soundcard output to triggers compatible with the ActiveTwo system.

3.1.4 Procedure.

The dataset contained evoked-response data recorded whilst participants listened to the story segments described in Sections 3.1.2. Upon completion of hearing assessment using PTA and a set of BKB sentences in quiet, participants were fitted with a 32 channel electrode cap conforming to the 10–20 system of electrode placement. After presentation of each segment participants were asked multiple-choice questions to confirm that they were attending and understood the speech. For recording of the evoked-responses participants were seated in a quiet room with a loudspeaker placed 1.2m directly in front.

3.1.5 Signal Processing.

3.1.5.1 Pre-processing.

Evoked response data was first re-referenced to the average. This data was then bandpass filtered into three bandwidths (delta, theta and wideband). For each band a one-pass, zero-phase, non-causal bandpass filter was created. The time-domain FIR windowing method with a hamming window was used for design of the filter. The resulting filter was a 6759 point FIR filter with a -6 dB cutoff at 0.5 Hz and 22.5 Hz for the wideband filter, 0.5 Hz and 5 Hz for the delta band, and 3 Hz and 9 Hz for the theta band. All filters had a passband ripple of 0.01 dB and a stopband attenuation of -53 dB. To decrease subsequent processing time, the resulting signals were then downsampled to 64 Hz using the Fourier method after windowing using a boxcar window in the frequency domain.

The absolute Hilbert envelope was computed to extract the speech envelope from the stimulus. This envelope was then filtered for each bandwidth using filters with identical properties to those used for the EEG data. The resulting filtered envelopes were then downsampled to 64 Hz to match the evoked-response data, again using an identical downsampling procedure. All data were preprocessed using the MNE Python package (Gramfort, 2013) and custom python scripts.

3.1.5.2 TRF Models.

As described in Section 2.2.2.5, the discrete-time TRF is a spatio-temporal model that maps the stimulus' speech envelope ($y(t_n)$) as a combination of evoked-response recordings $x_j(t+\tau)$ at a delay τ (note that $t+\tau$ is used to represent the non-causal mapping of the EEG data back to the amplitude envelope):

$$y(t) = \sum_{j=1}^{N} \sum_{\tau=1}^{T} \beta_{j,t} x_j(t+\tau) + \varepsilon_j(t), \tag{3.1}$$

where β is the weighting of evoked response channel j at time t (this can be thought of as a multi-channel impulse response, mapping the EEG channels to the envelope), T is the total number of delays, and $\varepsilon(t)$ is the residual of each channel that is not explained by the model. Note that $y(t)$ represents the amplitude envelope extracted directly from the digital representation of the audio file. This model can then be constructed by minimising mean squared error between the predicted stimulus reconstruction and the actual stimulus:

$$\min \varepsilon(t)^2 = \sum_{t=1}^{T} [y(t) - \hat{y}]^2 \tag{3.2}$$

Modelling the relationship between the evoked responses as a linear sum, β can be defined using the ordinary least squares approach:

$$\beta = (X'X)^{-1}(X'y), \tag{3.3}$$

where X is the design matrix and X' is its transpose. For calculation of $X'X$ in Equation 3.3 previous models have typically used either a Cholesky solver (Maddox and Lee, 2018; Gramfort, 2013) or a singular value decomposition solver (Crosse et al., 2016; Riecke et al., 2019; Vanheusden et al., 2020). The choice of an implementation's solver can influence the calculation of model parameter and so verification of model parity is important to confirm that results across the previous literature are comparable.

In practice, these models require regularisation to prevent overfitting of the model to the data used to construct it. L2 regularisation is commonly used (Crosse et al., 2016; Lalor et al., 2009; Vanheusden et al., 2020), adding a penalty term to the loss function as follows:

$$\min \varepsilon(t) = \sum_{t=1}^{T} [y(t) - \hat{y}]^2 + \lambda \sum_{j=1}^{N} \hat{\beta}_{j,t}^2, \tag{3.4}$$

where $\hat{\beta}$ is the fitted model and λ is the regularisation parameter. Higher values of λ results in greater L2 regularisation, which penalises large parameters in the model. This can result in less complex and therefore more generalisable models, which may increase predictive performance. For further information on ridge regression, refer to Bishop (2006).

In addition to the verification of the two previously proposed models (Cholesky ridge regression and SVD ridge regression), this study evaluates the performance of a third model: ElasticNet-net regression. ElasticNet regression expands on the previously described Ridge regression by combining both L1 (LASSO) and L2 (Ridge) regularisation in the loss function:

$$\min \varepsilon(t) = \sum_{t=1}^{T}[y(t) - \hat{y}]^2 + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{N} \hat{\beta}_{j,t}^2 + \alpha \sum_{j=1}^{N} |\hat{\beta}_{j,t}| \right), \tag{3.5}$$

where α is the regularisation parameter used for the relative influence of L1 and L2 regularisation terms.

The L1 regularisation added in ElasticNet provides a form of automatic evoked-response channel selection. Unlike L2 regularisation, L1 regularisation will give 0 weightings to channels/time-delays that do not contribute to the predictive power of a model, effectively removing them from the model. This may occur when channels contain artefacts or do not represent relevant neuronal activity. However L1 regularisation may perform poorly for channels that are strongly co-linear (as would be expected for evoked response recordings). The α regularisation parameter is provided to balance the use of regularisation types, allowing for combination with L2 regularisation which is well suited to highly correlated data. This model has been shown to provide improved predictive performance by combining the often superior performance of L2 regularisation (for problems where there are more observations (samples) than features (electrode channels)) with the feature selection capabilities of L1 regularisation (Zou and Hastie, 2005). The ElasticNet model used in the current study was fitted using coordinate descent, an iterative optimisation algorithm for efficiently finding the minimum of a loss function (Wright, 2015).

For each participant, models were fitted to the data using 5-fold cross-validation to evaluate their reconstruction performance. For all folds both stimulus and evoked-response data were standardised using z-scoring. For each fold an 80%–20% split was used, with 20 minutes of data used for training of the model and the remaining 5 minutes for testing. For each training fold, 50 λ values (logarithmically spaced between 10^{-15} to 10^{15}), and 10 α values (linearly spaced between 0.1 and 1.0) were used to construct a total of 500 models. For each test fold, the reconstructed envelope ($\hat{y}$) and the actual envelope (y) were split into 30 10-second segments. Pearson's correlation was calculated between y and $\hat{y}$ for each of the 30 segments. Optimal regularisation parameters were chosen as those which produced the highest average correlation coefficient across all test-fold segments. The mean and standard deviation of the correlation coefficients were then calculated from all test segments to provide a measure of model reconstruction performance. In addition, the null-distribution of the model was estimated. An identical procedure was performed as described above, but reversing the speech envelope provided to the model. This provided a measure of the model's chance level reconstruction performance.

All models were implemented in custom Python scripts, utilising the MNE package (Gramfort, 2013) and the Scikit-learn package (Pedregosa et al., 2011).

The data analysis detailed was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 52472). Data collection protocol was approved by the local National Health Service (NHS) ethics committee (refer to Vanheusden et al. (2020) for details). All research was performed in accordance with the relevant guidelines and regulations.

3.2 Statistics

Normality was assessed using a Shapiro-Wilk test. Data was not normally distributed within data splits across participants. Therefore non-parametric tests were used for subsequent analysis. Friedman tests were conducted for each bandwidth to assess the effect of model type. 6 planned post-hoc Wilcoxon signed-rank tests were performed for subsequent analysis (applying Bonferonni-Holm correction for multiple comparisons). Further unplanned post-hoc Wilcoxon signed-rank tests were conducted for exploratory analysis of individual participant results (no further correction was made for multiple comparisons at this stage). Statistics were calculated using custom Python scripts, utilising Scipy (Virtanen et al., 2020) and Pingouin (Vallat, 2018) packages.

3.3 Results

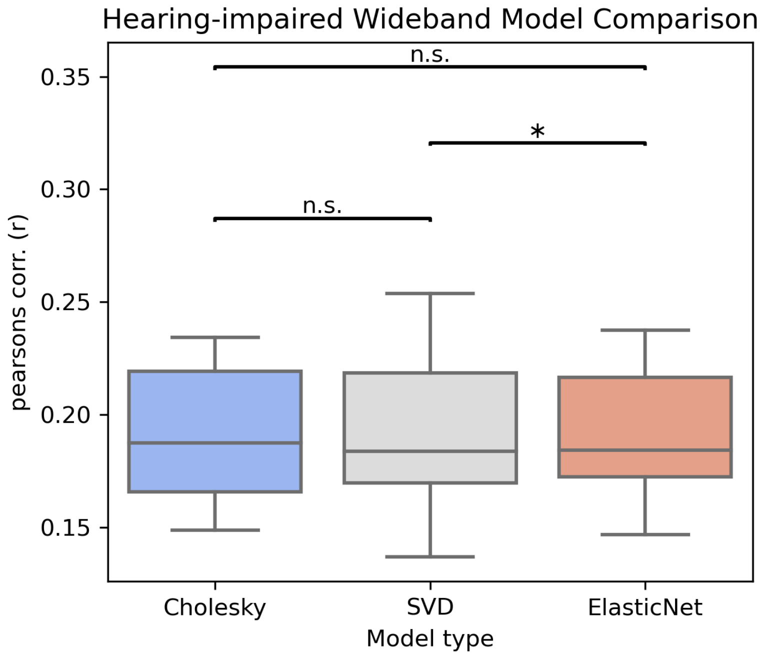

Figure 3.1: Group level hearing-impaired subject correlation for the wideband Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$* : p \leqslant 0.05$, n.s. : $p > 0.05$

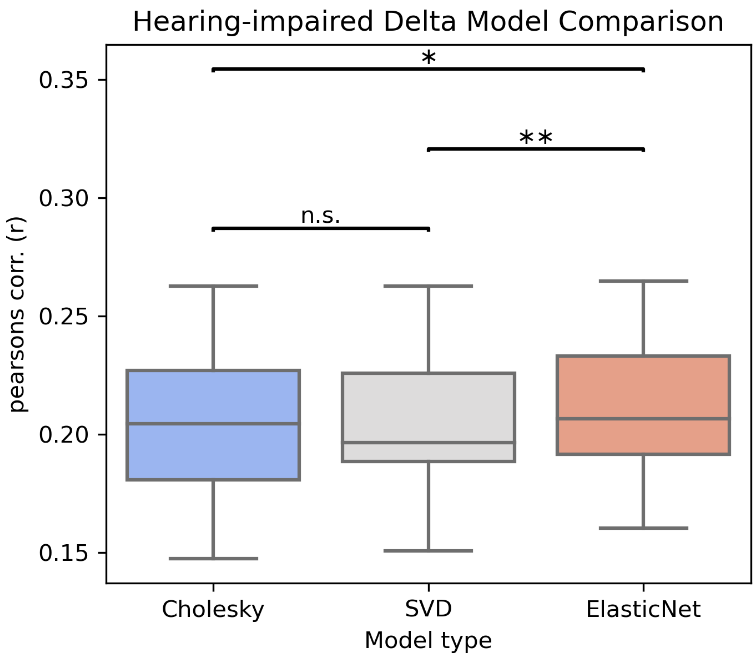

Figure 3.2: Group level hearing-impaired subject correlation for the delta Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

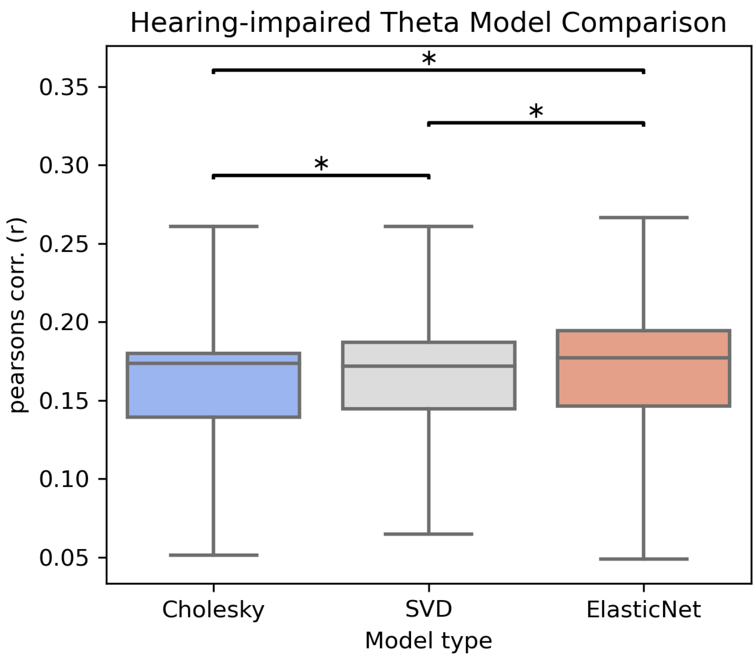

Figure 3.3: Group level hearing-impaired subject correlation for the theta Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$* : p \leqslant 0.05, \text{n.s.} : p > 0.05$

Friedman tests were conducted, correcting for ties, with model type (Cholesky ridge regression, SVD ridge regression and ElasticNet models) for the delta (0.5–4Hz), theta (4–8Hz) and wideband (1–20Hz) bandwidths. A significant effect of model type was found for wideband ($\chi^2(2) = 11.23, p = 0.004$) delta band ($\chi^2(2) = 16.62, p < 0.001$) and theta band ($\chi^2(2) = 7.54, p = 0.023$) conditions.

For the wideband condition, the median correlation for the Cholesky model was 0.187 (ranging from 0.149 to 0.234). Its time reversed model had a median correlation of -0.006 (ranging from -0.024 to 0.021). The median correlation for the SVD model was 0.184 (ranging from 0.137 to 0.254). Its time reversed model had a median correlation of 0.001 (ranging from -0.020 to 0.012). The ElasticNet model's median correlation was 0.184 (ranging from 0.147 to 0.237). Its time reversed model produced correlation of 0.0 for all measurements, due to the L1 regularisation.

For the delta band condition, the median correlation for the Cholesky model was 0.204 (ranging from 0.147 to 0.263). Its time reversed model had a median correlation of -0.006 (ranging from -0.028 to 0.021). The median correlation for the SVD model was 0.196 (ranging from 0.151 to 0.263). Its time reversed model had a median correlation of -0.001 (ranging from -0.032 to 0.016). The ElasticNet model's median correlation was 0.207 (ranging from 0.160 to 0.265). Again, the time-reversed ElasticNet produced correlation of 0.0 for all measurements, due to the L1 regularisation.

For the theta band condition, the median correlation for the Cholesky model was 0.173 (ranging from 0.051 to 0.261). Its time reversed model had a median correlation of -0.002 (ranging from -0.024 to 0.043). The median correlation for the SVD model was 0.172 (ranging from 0.065 to 0.261). Its time reversed model had a median correlation of 0.000 (ranging from -0.023 to 0.016). The ElasticNet model's median correlation was 0.177 (ranging from 0.049 to 0.266). As with previous conditions, the L1 regularisation resulted in correlations of 0.0 for the time-reversed ElasticNet.

6 planned post-hoc Wilcoxon sign-ranked tests were performed, correcting for multiple comparisons, to assess the difference in distributions between the Cholesky model, SVD model, and the ElasticNet model and to compare each model to its time-reversed equivalent model. For the wideband condition, correlations were significantly higher overall for the ElasticNet than the SVD model ($W = 9, p = 0.024$). No significant difference was found between the Cholesky model and the ElasticNet model ($W = 14, p = 0.053$) or SVD models ($W = 26, p = 0.191$). Time-reversed models performed significantly worse than all time aligned models (Cholesky model: $W = 0, p = 0.001$; SVD model: $W = 0, p = 0.001$; ElasticNet model: $W = 0, p = 0.001$). Results of the wideband model are illustrated in Figure 3.1. For the delta band condition, the ElasticNet model performed significantly better than both the Cholesky model ($W = 13, p = 0.043$) and the SVD model ($W = 0, p = 0.001$). The performance of the Cholesky and SVD models was not significantly different ($W = 44, p = 0.946$). Time-reversed models performed

significantly worse than all time aligned models (Cholesky model: $W = 0, p = 0.001$; SVD model: $W = 0, p = 0.001$; ElasticNet model: $W = 0, p = 0.001$). Results of the delta band model are illustrated in Figure 3.2. For the theta band, the ElasticNet model performed significantly better than both the Cholesky model ($W = 9, p = 0.024$) and the SVD model ($W = 10, p = 0.024$). The SVD model also performed significantly better than the Cholesky model ($W = 11, p = 0.024$). Time-reversed models performed significantly worse than all time aligned models (Cholesky model: $W = 0, p = 0.001$; SVD model: $W = 0, p = 0.001$; ElasticNet model: $W = 0, p = 0.001$). Results of the theta band are illustrated in Figure 3.3. 6 Further unplanned post-hoc wilcoxons sign-ranked tests were performed for each participant individually. These results are illustrated in Appendix A. No correction for multiple comparisons was applied for these exploratory tests. Significant differences are illustrated for model comparisons in Figures A.1, A.2 and A.3. All subjects were found to have strongly significant ($p > 0.001$) higher correlations for the time-aligned model than for the time-reversed model.

3.4 Discussion

In the current study, it was shown that the ElasticNet model provides a small but significant improvement when reconstructing the speech amplitude envelope for a range of hearing-impaired listeners. In all conditions ElasticNet performed significantly better than both the Cholesky and SVD models. However, the largest median correlation improvement shown for the theta band was only 0.004. In addition this study has verified the parity of previously proposed methods, finding small but significant differences in implementations for only the theta band (median correlation performance differed by only 0.001).

Despite the small magnitude of performance improvements presented, these results suggest that the ElasticNet may offer a suitable alternative, as it provides additional benefits over the Cholesky and SVD models. For example, the ElasticNet model produced a DC output for the reversed stimulus, demonstrating its ability to remove channels that do not contribute to predictions (thus resulting in a DC model when no channel contributes). This may have benefits when analysing noisy data, as L1 regularisation may provide a replacement for manual or set rule based removal of channels, (methods which may result in loss of useful data for the model). Additionally, as an L1 coefficient exists for each time delay, this may also provide further information on specific time-delays that contribute to the reconstruction. This information could be used in conjunction with direct analysis of the model's morphology to infer the underlying neural components that drive the stimulus reconstruction — analogous to traditional evoked-response analysis techniques (see Section 2.2.2 for details). Future work should look to further analyse the L1 coefficients to determine the channels and delays that best contribute to the model. This may provide insight into the number of channels that may be needed when moving towards a clinical implementation, and for faster testing times for current in-lab research using the TRF.

It was also unlikely that the ElasticNet model would perform worse than the other models as in situations where L1 regularisation is not effective, the α parameter allows the model to switch to only use L2 regularisation. Therefore, there is little disadvantage to using this model in place of previous models.

The comparison of the Cholesky and SVD methods showed no significant difference for wideband and delta band models. The difference between models for the theta band was small and unlikely to have any practical impact for comparison across studies. This suggests that these models are broadly interchangeable, with no discernible benefit of either method.

The limitations of this study should also be considered. It should be noted that the Theta band performed considerably worse than the other bands for all models. These results do not agree with previous literature using similar methodologies, where theta band reconstructions are typically greater than delta band reconstructions (Di Liberto et al., 2015; Lesenfants, 2019). This may be due to time alignment issues in the original dataset — it is possible that there is a delay/jittering of samples that may reduce reconstruction performance at this higher frequency band. However it is not possible to verify this with the available data. Therefore further analysis with an independent dataset may be necessary to verify these findings.

Furthermore, when interpreting the presented results other limitations of the dataset should be considered. This study demonstrated the efficacy of the presented methods for only hearing impaired listeners. The data was sufficient to provide a range of hearing losses for comparison of models, however further comparison may be necessary on a normal-hearing cohort. In addition, for this study the presented models were only compared for speech in quiet data. Further analysis is needed to assess the sensitivity of these models for speech-in-noise performance assessment. Studies using similar models have shown that these methods degrade in their reconstructive performance as noise increases, providing accurate speech-in-noise score predictions (Lesenfants, 2019; Decruy et al., 2018; Etard and Reichenbach, 2019). Additional exploration of responses to speech-in-noise is needed to determine the benefits of the ElasticNet model in particular for clinical or in-lab analysis of speech-in-noise performance.

There are a number of steps needed to maximise the potential of the TRF models for both clinical and in-lab applications. For example, in this study models were only fitted for typical cortical bandwidths (wideband, delta and theta) in the presented study. In order to analyse non-cortical components of hearing loss, the models presented could be adapted to reconstruct the temporal fine-structure of the stimulus, as demonstrated by Maddox and Lee (2018). This work would be necessary for the development of an analysis tool for assessment of the complete auditory pathways.

In this study the TRF models were fitted using the amplitude envelope as the stimulus. This stimulus feature may not be optimal for a linear decoder, given the non-linear nature of the auditory system. Further exploration of the potential benefits of non-linear perceptually motivated feature reconstruction is provided in Chapter 4. Performance of the ElasticNet model may also be improved using techniques such as variance inflation factor analysis (VIF) to assess and PCA to reduce collinearity in the input to the ElasticNet model. Collinearity in data is particularly detrimental to L1 regularisation so these collinearity reduction methods may provide better regularisation performance than is reported in this study (Dormann et al., 2013). Further work should also assess the performance of the ElasticNet model with reduced amounts of training data. Current research typically trains each model on 20 minutes of data, which is too long for these methods to be viable for clinical diagnostic or for intervention fitting procedures. With smaller datasets overfitting is more likely to occur and so the additional regularisation provided by ElasticNet may provide performance benefits that have not been shown in this study. The use of ElasticNet could also be extended to other areas, such as auditory attention decoding or for assessment of audio-haptic integration, following established methodologies, such as those presented by Riecke et al. (2019) and Fu et al. (2019). This may provide additional insight into the underlying neural mechanisms in these areas.

Overall, results indicate that any of these methods would be a viable option for analysis of evoked-responses to running speech for hearing impaired listeners. A new model has also been presented, which may offer further benefits for interpretation than traditional models used for assessment of evoked-responses to running-speech. This study indicates that the ElasticNet model has the potential for development as both a clinical and in-lab assessment of speech-in-noise performance.

3.5 Contributions

The data used in this study was originally collected for the study detailed in Vanheusden et al. (2020). This dataset was chosen as it was suitable for assessment of the signal processing methods proposed in Section 3.1.5.2, and for the diversity of participant responses within the dataset, as detailed in Section 3.1.1. All stimulus preprocessing and response analysis (detailed in Sections 3.1.5.1 and 3.1.5.2) were performed independently of original analysis described in Vanheusden et al. (2020). Reanalysis of this data was authorised by the original authors. Data analysis implementation was designed and implemented by the author Samuel Perry and supervisors Mark Fletcher, David Simpson and Steve Bell. The above work was published, in part in the conference proceeding by Perry et al., 2018, and Perry et al., 2019.

Chapter 4

Feature optimisation of the Temporal Response Function model as a speech-in-noise assessment method

The previous Chapter 3 outlined an adaptation of the TRF model as a method with potential to be used as a measure of speech-in-noise performance. This focused on the optimisation of a linear model for reconstructing the amplitude envelope of speech. However, the amplitude envelope is a basic acoustic property of the audio, and so does not represent the non-linear perception of amplitude in the auditory system. It may therefore be expected that a feature that better represents human perception of loudness would be better represented in the neural responses to running speech. Previous research suggests that perceptually motivated features such as spectrograms (Daube et al., 2019) and categorical phonetic features (Di Liberto et al., 2015; Di Liberto and Lalor, 2017) can improve the reconstruction performance of TRFs, and improve prediction of speech intelligibility (Lesenfants, 2019). In addition, a recent study by Biesmans et al., 2017 suggests that auditory models can improve the performance of the TRF for the task of auditory attention decoding. Their results show attention decoding accuracy improving from 77.7% for a standard envelope extraction method, to 81.5% with a simplistic loudness model. Furthermore, more complex loudness models are not found to provide further benefit in this study. Combined, these studies suggest that perceptual loudness TRF may improve the reconstruction performance. This may benefit the design of an evoked-response based measure of speech-in-noise performance, particularly if using speech-in-noise score prediction methods such as that presented by Vanthornhout et al., 2018.

The following Section 4.1 outlines a study to evaluate the TRF's reconstruction performance of a perceptual loudness model in normal-hearing and hearing impaired participants. In this chapter, the ElasticNet model proposed in previous Chapter 3 will be used to compare the reconstruction performance of the speech amplitude envelope and a perceptual model of loudness (ITU BS17704). The ITU loudness model produces a function of perceived loudness, for speech in background noise (International Telecomunication Union, 2015). It may therefore be expected that this will outperform the amplitude envelope in terms of reconstruction performance. It should be noted that the candidate models presented by Biesmans et al., 2017 may offer reasonable alternative models. However, the best performing model (the “p-law” model) does not benefit from the same level of validation as the ITU model presented: The ITU model was chosen as a simple (in terms of computational complexity, an important factor in the inclusion of the proposed methods in low-power/wearable devices, as highlighted by Biesmans et al., 2017 and discussed further in Section 4.4), standardized model that has been validated for 97 NHLs on a wide variety of material, from spoken words to sound effects and television shows. The model was validated using a subjective loudness matching procedure, where participants were tasked with matching various material (336 audio sequences) to a 60 dBA reference signal (a segment of female English speech). Subjective rating are highly correlated with the ITU model for these validation studies for 2 monophonic datasets (r=0.979 and r=0.985), and are also highly correlated for a further mixed monophonic and multi-channel dataset (r=0.980) (International Telecomunication Union, 2015). Therefore the ITU model appears a well suited candidate model for the proposed application. Implementation of this model is detailed in Section 4.1.5

The presented study evaluates the features on both normal-hearing and hearing impaired listeners, using a reduced amount of data-per participant in comparison to that used in Chapter 3. This aims to more comprehensively test the feasibility of the proposed method than that of Chapter 3 — using datasets of representative cohorts and EEG recording lengths that are closer to clinically feasible than in the previous chapter. As with the previous chapter, models will be assessed for 3 bandwidths – Wideband (0–20 Hz), Delta band (1–4 Hz) and Theta band (4–8 Hz). These bandwidths are thought to be in the range of amplitude envelope frequencies that contribute most to speech intelligibility (Aiken and Picton, 2008) and are commonly used for cortical TRF reconstruction. Results of this study are detailed in Sections 4.2 and 4.3. The potential for a loudness model based TRF to be used as a speech-in-noise score prediction method and the current limitations of the method as a diagnosis/intervention fitting procedure are outlined in Section 4.4.

4.1 Methods

4.1.1 Participants.

This study used evoked response data from two datasets: The dataset of evoked-responses from hearing-impaired listeners described in Chapter 3, and a dataset of responses from normal-hearing listeners. The normal-hearing data used in this study was originally collected by Stephanie Nel, a BSc student at the University of Southampton. The dataset contained evoked response data for 13 normal-hearing participants (5 males, 8 females, aged between 18 and 29). Participants were native English-speaking and had normal hearing, as assessed using PTA.

4.1.2 Stimuli.

For the normal-hearing dataset, 12 minutes of running speech was used for stimulation of evoked responses. The stimulus was taken from a freely available audiobook (Colum, 2021) read by a female speaker. The audiobook was split into 4 segments of approximately 3 minutes each. The speech was low-pass filtered at 3 KHz with a 120th order FIR filter. The stimulus was normalised to a peak of 0 dB FS. Stimulus was presented at 75 dBA equivalent sound pressure level (LeqA SPL). For the normal hearing dataset, participants were presented this stimulus as part of a wider battery of stimuli. Recordings of the other stimuli were not used as part of this research.

4.1.3 Apparatus.

For the normal-hearing dataset, the stimulus was presented to the participant via a loudspeaker placed 1.2 m in front of the participant. The loudspeaker was driven by an RME BabyFace soundcard (RME, n.d.). EEG data were collected using a BioSemi ActiveTwo EEG system (BioSemi, n.d.). Thirty-two standard electrodes were used, arranged following the 10–20 electrode system layout. Evoked response data was recorded at a samplerate of 2048 Hz. Triggers were used to synchronise stimulus output and evoked-response recordings. These were produced via the soundcard, using an Arduino Due (Arduino, 2021) based trigger system with custom software to convert the soundcard output to triggers compatible with the ActiveTwo system.

4.1.4 Procedure.

The procedure for collection of data for both the hearing-impaired dataset and the normal hearing dataset was similar, with the exception of the location, tester and test procedure (as noted where applicable). The location of data collection for the normal-hearing dataset was the Hearing and Balance Centre at the University of Southampton. For the hearing-impaired dataset, data was collected in the Royal Berkshire NHS Foundation Trust Audiology Department.

4.1.5 Signal processing

Pre-processing of both the evoked-response data and amplitude envelope were performed as detailed in previous Section 3.1.5.1. In addition, the loudness stimulus feature was extracted using the procedure illustrated in Figure 4.1, following the specification detailed in ITU BS.1770–4 (International Telecomunication Union, 2015).

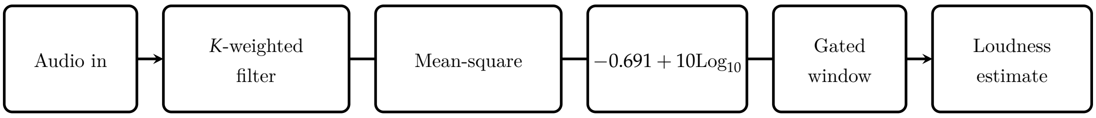

Figure 4.1: Schematic diagram of the ITU BS.1770–4 loudness model

First a K–weighted filter was applied. This consisted of a high-shelf filter and a high-pass filter, implemented as two IIR filters in series. Filter morphology is illustrated in detail in International Telecomunication Union (2015). The mean-square of each sample is then calculated and converted to the $LKFS$ scale, defined as

$$l(t) = -0.691 + 10\text{Log}_{10}(z(t)), \tag{4.1}$$

where z is the mean-square of the input sample at time t. The constant of 0.691 is subtracted to cancel the additional gain applied to the signal by the K-weighted filter. The signal is then split into 400 ms windows, with an overlap of 75%. Each window is then gated using an absolute and a relative threshold. The absolute threshold (Γ_A) is calculated as

$$\Gamma_A = 10^{(-70.0+0.691)/10}. \tag{4.2}$$

Samples that are below this threshold are first removed from further calculations. The relative threshold (Γ_R) is then defined as

$$\Gamma_R = -0.691 + 10\text{Log}_{10}\left(\sum_J z\right) + 10^{(-10.0+0.691)/10}, \tag{4.3}$$

where $J = \{t : z > \Gamma_A\}$. The loudness of each block of samples is then calculated as:

$$L = -0.691 + 10\text{Log}_{10}\left(\frac{1}{|K|}\sum_K z\right) \tag{4.4}$$

where L is the loudness measure of the current block, $K = \{t : z > \Gamma_A \text{ and } z > \Gamma_R\}$ and $|K|$ is the number of elements in K. The resulting momentary loudness measure was then filtered into each of the analysis bands (Delta, Theta and Wideband) using an identical filter to the amplitude envelope, as specified in Section 3.1.5.1. An illustration of the ITU loudness model's output in relation to a standard amplitude envelope is shown in Figure 4.2.

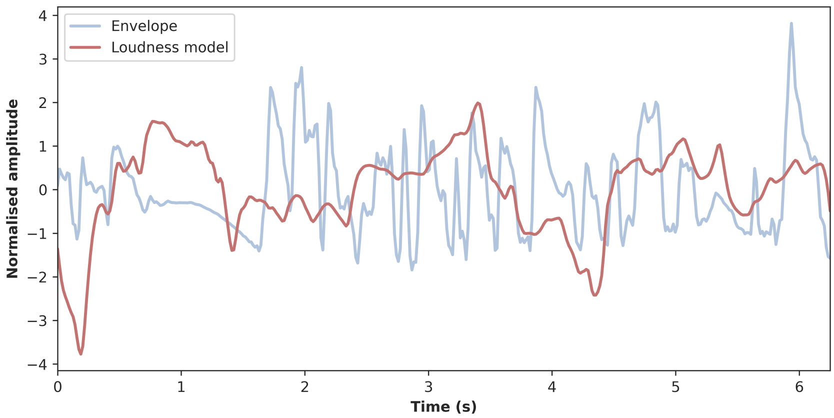

Figure 4.2: Comparison of the normalised amplitude envelope and ITU loudness model function for a segment of stimulus

The ElasticNet model defined in previous Chapter 3.1.5.2 was used for reconstruction of the stimulus features. For each dataset, two models were optimised for each participant. The first was fitted using the speech amplitude envelope of the stimulus, and the second was fitted using the momentary loudness measure. An example comparison of an original amplitude envelope to its reconstruction is illustrated in Figure 4.3. 5-fold cross-validation was used, with an 80%–20% train–test split per fold. This resulted in approximately 9 minutes 40 seconds of training data and 2 minutes 20 seconds of test data per fold. For each fold, the input feature and evoked response data were standardized using z-scoring. Models were optimised per training fold using a grid search to select the best of 50 λ parameter values (ranging logarithmically between 10^{-15} and 10^{15}) and 10 α parameter values (ranging linearly between 0.0 and 1.0). For details on model parameters, please refer to Chapter 3. For the test data, feature reconstructions ($\hat{y}_{\text{env}}$ and $\hat{y}_{\text{ITU}}$) and their respective targets (y_{env} and y_{ITU}) were split into 12 10-second segments. The Pearson's correlation was calculated for each segment pair. Averaging these correlations provided a measure and standard-deviation of feature reconstruction performance. Mean correlation was used as the performance metric in the grid search for the optimal model parameters. Null-distributions were calculated for the final models of each feature. This was achieved by fitting two further models, using time reversed features to provide a measure of chance-level reconstruction performance.

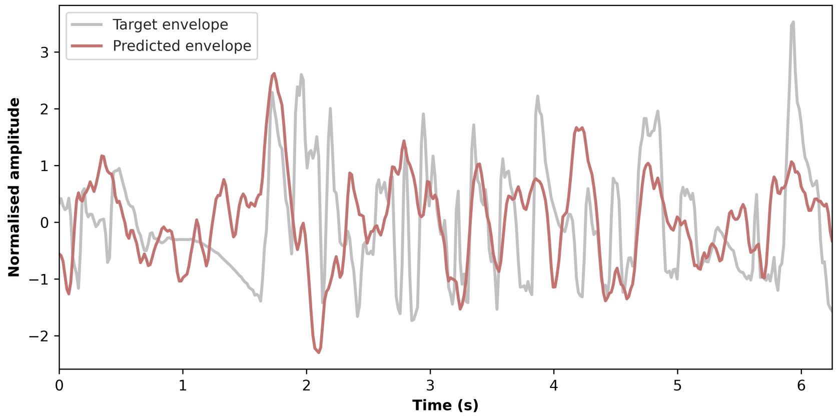

Figure 4.3: Comparison of the original and reconstructed amplitude envelope, reconstructed using the ElasticNet model, for a segment of stimulus

All data were preprocessed using the MNE Python package (Gramfort, 2013) and custom python scripts.

The data analysis detailed was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 52472). Hearing impaired participant data collection protocol was approved by the local National Health Service (NHS) ethics committee (refer to Vanheusden et al. (2020) for details). Normal-hearing participant data collection protocol was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 17552). All research was performed in accordance with the relevant guidelines and regulations.

4.2 Statistics

For both the normal-hearing and hearing-impaired dataset, normality was assessed using a Shapiro-Wilk test. Data was not normally distributed within data splits across each participant. Therefore non-parametric tests were used for subsequent analysis. 3 Friedman tests were conducted for each dataset (using Bonferonni-Holm correction for multiple comparisons), with feature type as the primary factor. Planned Wilcoxon signed-rank tests were performed for all models that yielded a significant results from the Friedman test (correcting for multiple comparisons using Bonferroni-Holm correction for multiple comparisons). Further unplanned Wilcoxon sign-ranked tests were performed as exploratory analysis per-participant. For these tests, no further correction was made for multiple comparisons. Statistics were calculated using custom Python scripts, utilising Scipy (Virtanen et al., 2020) and Pingouin (Vallat, 2018) packages.

4.3 Results

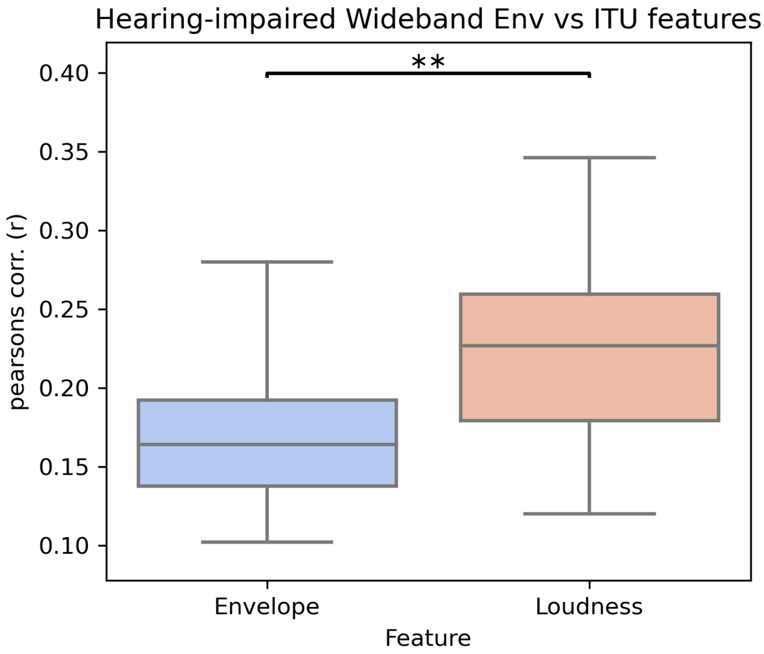

Figure 4.4: Group level hearing-impaired participant feature reconstruction correlation for the wideband band amplitude envelope and perceptual loudness models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $** : p \leqslant 0.01$

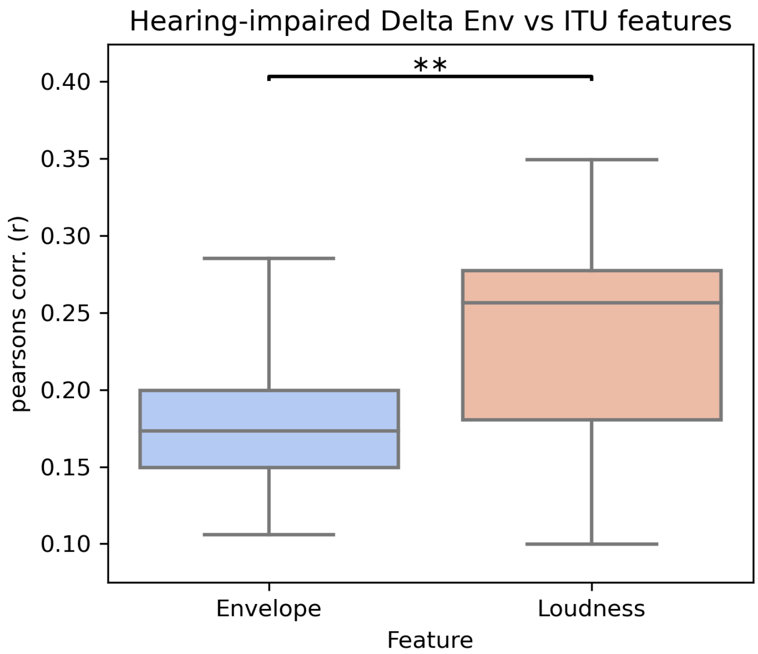

Figure 4.5: Group level hearing-impaired participant feature reconstruction correlation for the delta band amplitude envelope and perceptual loudness models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $** : p \leqslant 0.01$

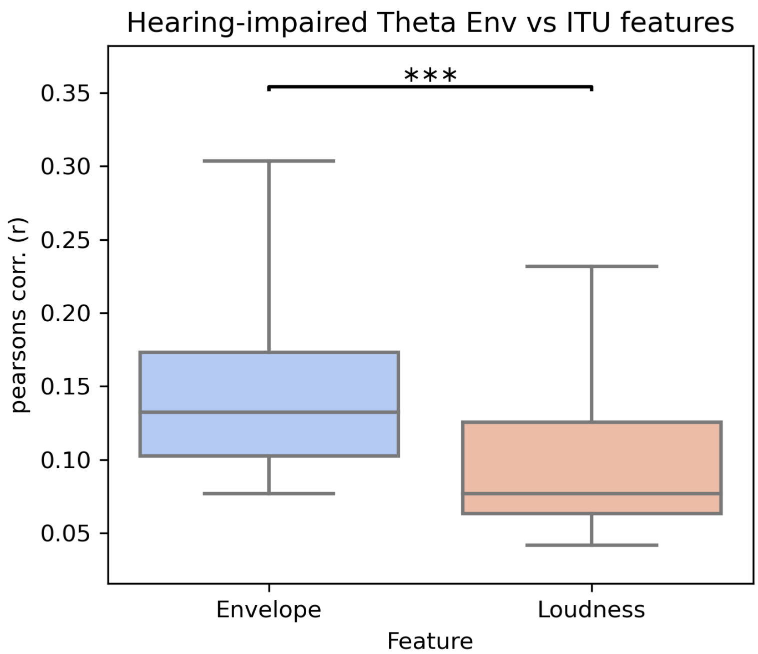

Figure 4.6: Group level hearing-impaired participant feature reconstruction correlation for the theta band amplitude envelope and perceptual loudness models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $*** : p \leqslant 0.001$

Separate analysis was performed for the normal-hearing and hearing impaired datasets, due to discrepancies in data collection methods that limited comparison (such as differences in testers and testing protocols). The rationale for this is discussed in Section 4.4.

For the hearing-impaired dataset, Friedman tests were performed, correcting for ties, with envelope and loudness features as factors. Results showed a significant effect of feature type for wideband ($\chi^2(1) = 5.4, p = 0.02$), delta band ($\chi^2(1) = 8.07, p = 0.005$) and theta band ($\chi^2(1) = 11.27, p < 0.001$).

For the wideband condition, the median correlation for the amplitude envelope was 0.164 (ranging from 0.147 to 0.280). Its time reversed model had a median correlation of 0.009 (ranging from -0.036 to 0.045). The median correlation for the loudness model was 0.227 (ranging from 0.120 to 0.346). Its time reversed model had a median correlation of -0.014 (ranging from -0.048 to 0.048)

For the delta band condition, the median correlation for the amplitude envelope was 0.173 (ranging from 0.106 to 0.285). Its time reversed model had a median correlation of -0.001 (ranging from -0.049 to 0.054). The median correlation for the loudness model was 0.256 (ranging from 0.100 to 0.349). The time reversed model had a median correlation of -0.006 (ranging from -0.049 to 0.064)

For the theta band condition, the median correlation for the amplitude envelope was 0.133 (ranging from 0.077 to 0.303). Its time reversed model had a median correlation of 0.008 (ranging from -0.045 to 0.04). The median correlation for the loudness model was 0.077 (ranging from 0.042 to 0.232). The time reversed model had a median correlation of -0.002 (ranging from -0.073 to 0.041)

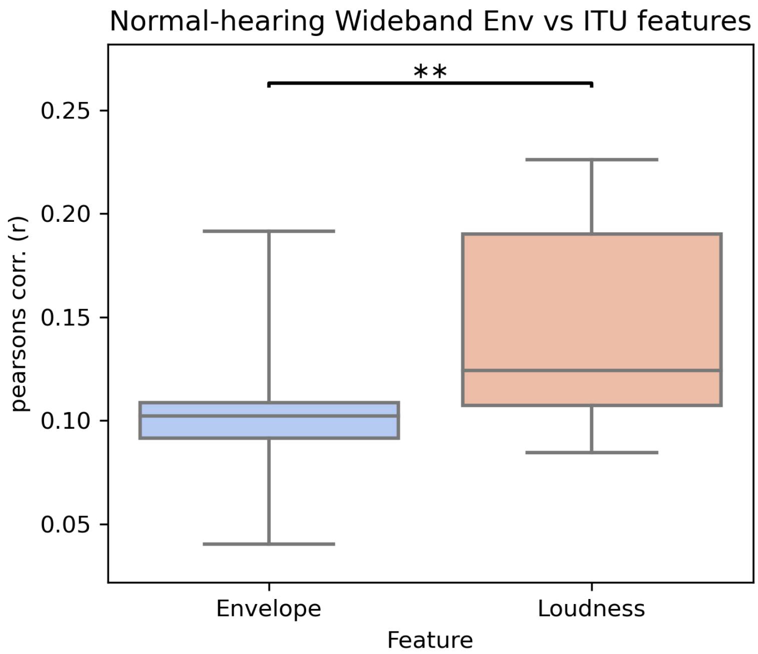

Figure 4.7: Group level normal-hearing participant feature reconstruction correlation for the wideband amplitude envelope and perceptual loudness models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $** : p \leqslant 0.01$.

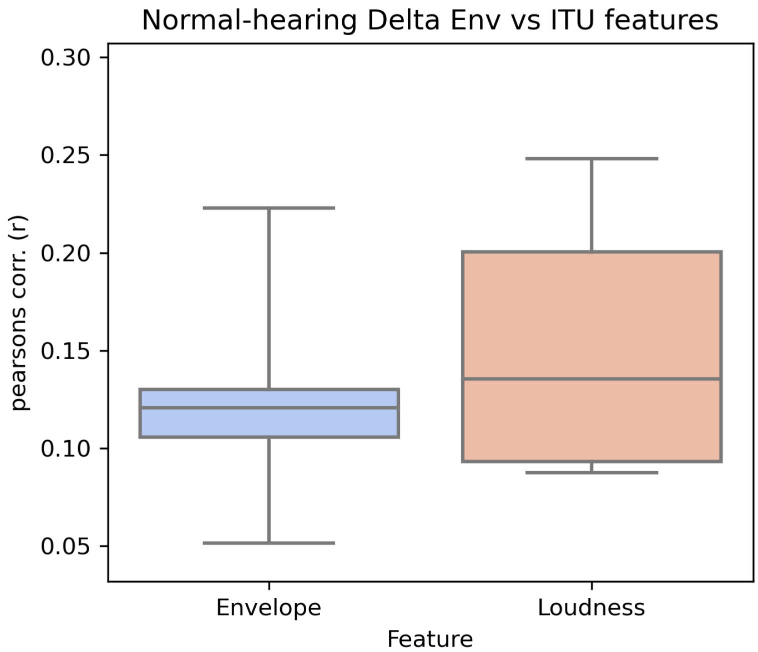

Figure 4.8: Group level normal-hearing participant feature reconstruction correlation for the delta band amplitude envelope and perceptual loudness models.

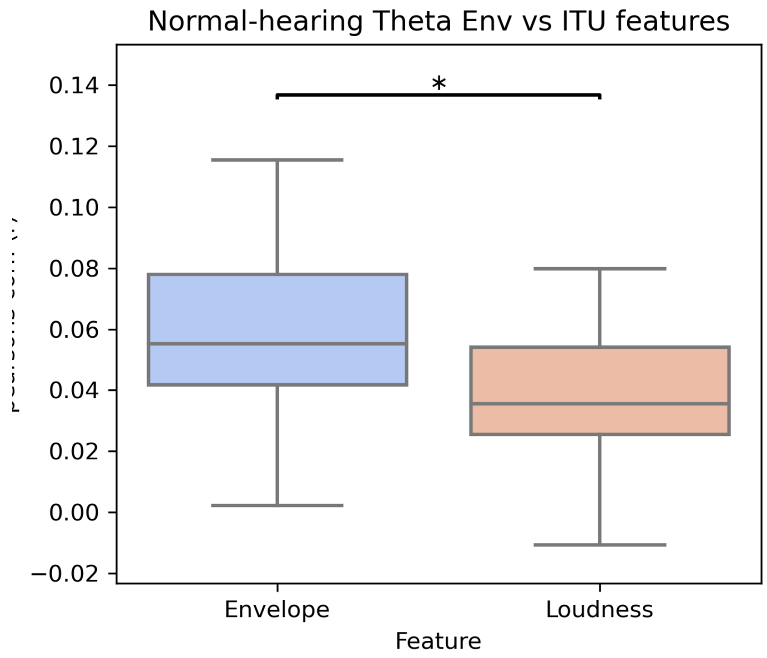

Figure 4.9: Group level normal-hearing participant feature reconstruction correlation for the theta band amplitude envelope and perceptual loudness models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $* : p \leqslant 0.05$.

For the normal-hearing dataset, an effect of feature type for both wideband ($\chi^2(1) = 6.23, p = 0.01$) and theta ($\chi^2(1) = 11.27, p < 0.001$) band was found. For the delta band ($\chi^2(1) = 3.77, p = 0.052$) results were approaching, but did not meet the threshold for significance.

For the wideband condition, the median correlation for the amplitude envelope was 0.164 (ranging from 0.147 to 0.280). Its time reversed model had a median correlation of 0.009 (ranging from -0.036 to 0.045). The median correlation for the loudness model was 0.227 (ranging from 0.120 to 0.346). Its time reversed model had a median correlation of -0.014 (ranging from -0.048 to 0.048)

For the delta band condition, the median correlation for the amplitude envelope was 0.173 (ranging from 0.106 to 0.285). Its time reversed model had a median correlation of -0.001 (ranging from -0.049 to 0.054). The median correlation for the loudness model was 0.256 (ranging from 0.100 to 0.349). The time reversed model had a median correlation of -0.006 (ranging from -0.049 to 0.064)

For the theta band condition, the median correlation for the amplitude envelope was 0.133 (ranging from 0.077 to 0.303). Its time reversed model had a median correlation of 0.008 (ranging from -0.045 to 0.04). The median correlation for the loudness model was 0.077 (ranging from 0.042 to 0.232). The time reversed model had a median correlation of -0.002 (ranging from -0.073 to 0.041)

In each dataset, 4 planned post-hoc Wilcoxon sign-ranked tests were performed per band (excluding the normal-hearing delta band, which did not reach significance in the Friedman test) for each participant. These assess the difference in reconstruction performance of the ElasticNet models for the speech-amplitude envelope and the perceptual loudness model, and also compared each model to its time-reversed model.

For the hearing-impaired dataset, the wideband models performed significantly better for the loudness model than for the speech-amplitude envelope ($W = 8, p = 0.002$). Both models performed significantly better than their time-reversed equivalents (amplitude envelope model: $W = 0, p = 0.000$; loudness model: $W = 0, p = 0.000$). The difference between features are illustrated in Figure 4.4.

For the delta band, the loudness model also significantly outperformed the envelope based model ($W = 7, p = 0.001$). Again, both models outperformed their respective time reverse models (amplitude envelope model: $W = 0, p = 0.000$; loudness model: $W = 0, p = 0.000$). Comparison of feature models for the delta band is illustrated in Figure 4.5. For the theta band, the loudness model was significantly worse that the amplitude envelope model ($W = 1, p = 0.000$). As with the previous models, both models outperformed their time-reversed models (amplitude envelope model: $W = 0, p = 0.000$; loudness model: $W = 0, p = 0.000$). Model comparison is illustrated in Figure 4.6.

For the normal-hearing dataset, the wideband models performed significantly better for the loudness model than for the speech-amplitude envelope ($W = 8, p = 0.002$). Both models performed significantly better than their time-reversed equivalents (amplitude envelope model: $W = 0, p = 0.000$; loudness model: $W = 0, p = 0.000$). The difference between features are illustrated in Figure 4.7. However, after correcting for multiple comparisons, the delta band loudness model did not significantly outperform the amplitude envelope model based on the Friedman tests detailed above. Therefore no further group-level analysis was performed of this band. Results for each feature model are illustrated in Figure 4.8.

For the theta band, the amplitude envelope model performed significantly better than the loudness feature model for the normal-hearing dataset ($W = 15, p = 0.017$). All models performed significantly better than their time-reversed equivalents (amplitude envelope model: $W = 1, p = 0.001$; loudness model: $W = 5, p = 0.002$). Results for the theta band are illustrated in Figure 4.9.

In addition to group level analysis, 3 unplanned post-hoc Wilcoxons sign-ranked tests were performed for each participant. These tests compared the performance of the amplitude envelope model to the loudness model, and also compared each to its time-reversed model. No correction for multiple comparisons was applied for these exploratory tests. Results of these tests are presented in Appendix B.

4.4 Discussion

Overall, this study has shown that for the delta and wideband TRFs, the loudness model can be reconstructed significantly better from the EEG signal than the amplitude envelope for both hearing-impaired and normal hearing listeners. Median reconstruction correlations (r) increased by as much as 28.5% for the hearing-impaired dataset and 31.1% for the normal-hearing dataset using the loudness model. These models were significantly above the noise floor at the group level. In addition only a minority of per-participant reconstructions were not significant (further analysis is required to determine the cause of these poor reconstructions). However, for the theta band, the loudness model consistently performed poorly compared to the amplitude envelope. The amplitude envelope median reconstruction performance was as much as 41.0% higher for the hearing-impaired dataset and 34.0% for the normal-hearing dataset. In addition, although comparison at group level showed significance of loudness model for the theta band above the time-reversed equivalent model (assessing the median reconstruction values for each participant for this analysis), measures within participant suggest performance wasn't significantly above the time-reversed model for the majority of participants.

The presented findings show consistent reconstruction improvements in the delta and wideband models. This suggests that a non-linear perceptual correlate of the amplitude envelope is well represented in the evoked responses to clean running speech. Further work is needed to asses the degradation of these reconstructions with added noise, and the correlation between this and subjective measures of speech-in-noise performance.

The theta models performed poorly for the loudness model. A possible reason for this was the 400 ms windowing in the loudness model. The 400 ms mean-square produces a low-pass moving average filter which may have filtered the higher frequency stimulus fluctuations of the theta band. Adjustment of the window parameter of the loudness model may result in more robust reconstructions for this bandwidth. Using the ITU specification loudness model this bandwidth's reconstruction performance will be limited. Adjustments to the window size may address this issue, but this modification may limit the model's correlation with behavioural measures of loudness perception.

It should also be noted that the variance of loudness model is consistently higher than the amplitude envelope. This may suggest that that the model is more sensitive to differences in stimulus, or that the loudness feature reconstruction is more susceptible to irrelevant interfering noise. In the former case, this may be advantageous for assessment of speech in noise performance, as it indicates that the feature is more sensitive to the stimulus, and so may be more sensitive to the effects of noise on speech intelligibility. This is supported by the results of Biesmans et al. (2017) which suggest that loudness models are more sensitive to competing talkers than amplitude envelope based models.

A limitation of the current study is that the normal-hearing dataset cannot be directly compared to the performance of the hearing impaired dataset, due to differences in collection protocols (such as differing room acoustics and speech presentation level). Future research may address this issue by comparing these cohorts with a unified data collection protocol. In the pursuit of a clinically viable metric of speech-in-noise performance, the amount of stimulus presented, as well as the number of electrodes used for the collection of data should be considered, given the limited time available for assessment. Methods such as sequential testing (Chesnaye et al., 2019) may be adapted to determine the optimal quantity of evoked-responses needed for fitting of loudness model TRFs in clinic. Further analysis of the model and L1 coefficients may also provide insight into the most prominent electrodes for analysis in clinic, as discussed in Section 3.4. It should also be considered that there was an unavoidable differences in the age of cohorts for the datasets of the current study. Future work should explore the performance of loudness model based TRFs for matched cohorts of hearing impaired and normal-hearing listeners. Finally, the effect of participant interventions has also not been assessed in the current study. Given that previous studies have shown no clear effect of hearing aids on amplitude envelope responses (Vanheusden et al., 2020), the effect of an intervention used to improve speech in noise performance should be assessed to understand the sensitivity of the TRF to the perception of audio provided by these devices.

There are several further steps required to maximise the impact of the proposed method. For the development of a speech-in-noise score prediction method, the presented regression model/loudness feature should be adapted, using a technique such as that presented by Vanthornhout et al. (2018) or Lesenfants (2019), to assess the sensitivity of the method to variations in background noise, and to understand how this correlates with subjective measures of speech-in-noise performance. This study suggests that providing perceptual loudness features can provide performance reconstruction improvements. Further augmentation of the proposed feature, adding additional categorical phonetic data may provide further improvements to reconstruction performance, and may provide additional information on higher level cortical processing, as discussed using similar methodology by Di Liberto et al. (2015) and Lesenfants (2019).

It may also be possible to increase reconstruction performance of the model by using non-linear regression models (in place of the ElasticNet model) in addition to perceptually motivated features. Recent studies have shown improved reconstruction performance for the amplitude envelope using dilated convolutional neural networks (Accou et al., 2021), as well as standard deep neural network architectures for attentional decoding (de Taillez et al., 2020) and for envelope reconstruction with intracranial EEG (Akbari et al., 2019; Yang et al., 2015). The ElasticNet model could also be further optimised by adapting it as a Linear Support Vector Machine (SVM) to allow for increased computational efficiency (Zhou et al., 2014). A simple adaption of this SVM model to use a non-linear kernel may provide an alternative method for introducing a data-driven non-linearity to the TRF model, in addition to the non-linearity introduced by the presented loudness model feature.

Finally, this method may have further application as an assessment method for cochlear implant users. This may be of particular use for objective and automatic fitting of cochlear implants, and may have application in closed loop cochlear implants that can re-tune to maximise speech-in-noise performance on the fly (Geirnaert et al., 2021). Previous work has suggested that the method could be readily translated for use in cochlear implant users (Verschueren et al., 2019). Further disucssion on the use of these methods in the context of assessing and optimising haptic interventions for cochlear implant users is also provided in Chapter 7

In summary, the presented results suggest that a loudness model based TRF method can provide significantly improved stimulus reconstruction performance for both normal-hearing and hearing-impaired listeners. Appropriate analysis bands have also been identified for the presented loudness model feature. The method presented may have further application as a predictor of speech-in-noise scores, and may have further applications such as analysis of cochlear implant users and in the development of neuro-steered hearing interventions, Further work is needed to assess the method's sensitivity to speech in noise stimulus and adaption of the method is necessary to produce a practical clinical measure.

4.5 Contributions

In addition to data used in previous Chapter 3 (see Section 3.5 for details), The data used in this study was originally collected by Stephanie Nel in completion of her BSc in Audiology at The University of Southampton. This dataset was chosen as it was suitable for assessment of the signal processing methods proposed in Section 4.1.5. All stimulus preprocessing and response analysis (detailed in Section 4.1) were performed independently of original analysis. Reanalysis of this data was authorised by the original authors. Data analysis implementation was designed and implemented by the author Samuel Perry and supervisors Mark Fletcher, David Simpson and Steve Bell. The above work was published, in part in the conference proceeding by Perry et al., 2018, and Perry et al., 2019.

Chapter 5

A haptic neuroprosthetic to improve pitch discrimination performance in cochlear implant listeners

As highlighted in Section 2.1, pitch is an important feature of speech, used to identify accent, age and sex (Abberton and Fourcin, 1978; Titze, 1989), as well as segmental (Oxenham, 2008; David et al., 2017), and suprasegmental properties (Banse and Scherer, 1996; Murray and Arnott, 1993; Most and Peled, 2007; Peng et al., 2008; Meister et al., 2009; Xin Luo et al., 2007). As detailed in Section 2.3.2, cochlear implant users are particularly poor at perceiving fine-grained pitch cues, due to the limited resolution and distortions introduced by the implant. When discriminating the difference in the pitch of two stimuli, cochlear implant user performance varies markedly. Studies report discrimination thresholds of as high as 80–90% pitch differences (roughly 10–11 semitones) required for musical instruments (Brockmeier et al., 2011; Bruns et al., 2016), and of around 10–20% (around 2–4 semitones) for synthetic tone complexes (Drennan et al., 2015; Kang et al., 2009). In each study, the variance in performance is high, with some users scoring as poorly as more than 100% (over an octave) to, as well as 0.5% (less than one semitone).

By providing additional pitch cues, a suitable haptic augmentation may be able to improve on the currently limited performance of cochlear implant users. As discussed in Section 2.4.2.2, one study by Huang et al. (2017) explores using a pitch-to-haptic mapping to improve speech-in-noise performance. Results showed promise, with improvements of 2.2dB SRT on average. However, to the author's knowledge, no studies have explored the limits of pitch discrimination possible when providing these cues synchronously via haptics with or without a cochlear implant. In addition, the work presented by Huang et al. (2017) extracted the haptic signal from clean speech, and therefore did not assess the noise-robustness of the signal-processing strategy. This factor that could inhibit the real-world application of a pitch based haptic stimulation method.

The presented study assesses the potential for haptic stimulation to improve pitch discrimination performance. Section 5.1 will first outline the design and implementation of a forearm worn device to provide haptic stimulation to enhance pitch discrimination performance, and its accompanying signal-processing strategy. The presented device design aims to be readily translatable to an inexpensive real-world intervention. A study is then detailed, with the following aims:

- Test the limits of pitch discrimination for simulated cochlear implant users with the designed device.
- Assess the robustness of the device and accompanying signal-processing strategy to inharmonic background noise
- Test whether the audio and haptic stimulation is effectively combined by the listener, to provide better performance than either modality alone

Results are then presented and discussed in Sections 5.2 to 5.4, outlining contributions made to the development of a real-world intervention for improving speech-in-noise performance in cochlear implant users. The limitations of the implementation are also discussed highlighting areas that may increase the performance and viability of the outlined device.

5.1 Methods

5.1.1 Device design

The aim was to create an in-lab device that would be capable of providing haptic stimulation up the length of the forearm. This would allow for a spatial mapping of F_0 estimates (the acoustic correlate of pitch) to locations up the arm, avoiding issues such as the poor frequency resolution of the tactile system and the lack of inexpensive motors with independent frequency and amplitude control, as highlighted in Section 2.4. Further, the design was limited to use technologies that could reasonably be adapted to a real-world, inexpensive intervention (such as inexpensive motors and motor drivers). An image of the device during development is provided in Figure 5.1

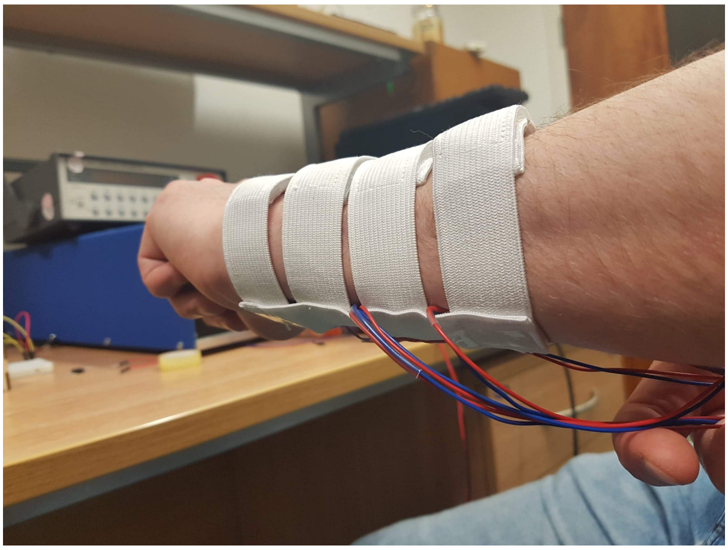

Figure 5.1: Image of an 8 motor prototype of the presented device during development. Each band hold 2 motors, on on the dorsal side and one on the palmar side of the wrist.

Of the motors available (detailed in Section 2.4.2), two ERM motors were selected: The Precision Microdrives 30610H (labelled as “Motor Type 1” in Figure 5.2) and 304116 (labelled as “Motor Type 2” in Figure 5.2) (Precision Microdrives, 2021a; Precision Microdrives, 2021b). These motors provided a suitable dynamic range (with maximum outputs of 1.84 G and 1 G respectively) and had differing frequency responses (operating frequency of 230 Hz for the 30610H and 280 Hz for the 306116). This allowed for an interleaved design with motors alternated along the forearm. This approach was taken to increase discrimination between motor locations, allowing users to discriminate motors both in space and in frequency. A total of 12 interleaved motors were used to match the output of the signal-processing strategy described in the following Section 5.1.2. Six motors were mounted on the dorsal (upper) side of the forearm and 6 on the palmar (lower) side, with motors spaced 3 cm apart. This spacing was selected to be greater than 2 point-discrimination thresholds for the forearm(Cholewiak and Collins, 2003; Schatzle et al., n.d.; Lévêque et al., 2000). An illustration of the motor arrangement is provided in Figure 5.2.

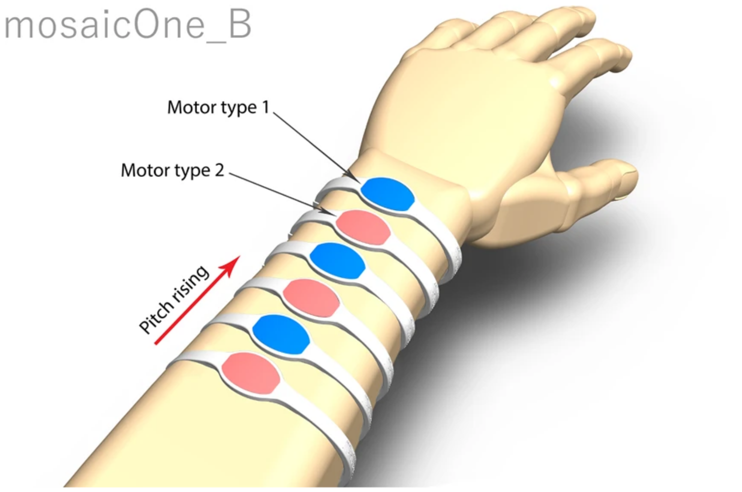

Figure 5.2: Schematic representation of the haptic device on the forearm. The two interleaved motor types are represented by different colours[1].

Motors were driven using custom made MOSFET based amplifiers. These amplifiers took control voltages as input and output driving voltage from 0 V to the rated voltage of the respective motors. Control voltages were generated using a DC-coupled MOTU 24Ao (MOTU, 2019), controlled by a custom Max (Cycling 74, 2019) patch (detailed in the following Sections 5.1.2 and 5.1.7). Motors were calibrated to rise linearly between the extrema of their respective outputs, to maximise their dynamic range.

5.1.2 Signal processing

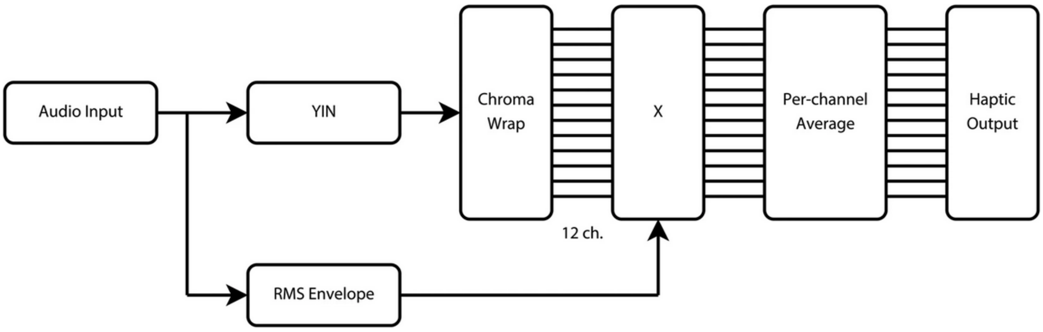

Figure 5.3: Schematic illustration of the signal processing strategy used to map input audio to haptic stimulation[1].

The proposed signal-processing strategy was designed to provide sub-octave pitch contours that are typically poorly perceived by cochlear implant users. To achieve this a

[1]Material from: 'Fletcher et al., Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic, Scientific Reports, published 2020, Springer Nature Limited

[1]Material from: 'Fletcher et al., Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic, Scientific Reports, published 2020, Springer Nature Limited

chroma-wrapped F_0 estimator was used, which categorises the estimated frequency as one of 12 semitone pitches within an octave. This produces a relative representation of pitch, presenting frequency changes as small as a semitone, whilst removing absolute pitch height information. The signal processing strategy used to map audio to tactile stimulation is illustrated in Fig 5.3. To generate the haptic signal, the F_0 of the audio was first estimated using the YIN fundamental frequency estimator, implemented in the Max Sound Box toolbox (IRCAM, 2018). A window size of 14 ms was used with no downsampling (giving a minimum possible estimate of 70 Hz). This F_0 estimate was then used to activate one of the 12 motors. To map the frequency estimate to a motor, the F_0 was first converted to the MIDI scale, a scale used to map frequency to musical pitch. The full mapping from estimated frequency to motor was calculated as:

$$F_{wrap}[n] = \mathrm{mod}\left(69 + 12 \cdot \log_2\left(\frac{F_0[n]}{440}\right), 12\right), \tag{5.1}$$

$$y_i[n] = \begin{cases} 1, & i = F_{wrap}[n] \\ 0, & \text{otherwise} \end{cases}, \tag{5.2}$$

where F_{wrap} is an integer in the range $0 \leq F_{wrap} < 12$, and y_i is the channel at index i. Using this strategy, it was expected that a minimum F_0 discrimination performance of around 6% (1 semi-tone) would be achievable (the minimum difference in F_0 required to switch between motors). If achieved, this performance would be markedly less than the 10–20% pitch difference discrimination performance of cochlear implant users shown in previous studies (Drennan et al., 2015; Kang et al., 2009). In parallel, the RMS amplitude was calculated using 14 ms windows. This envelope was then used to modulated the activated motor channel. Initial evaluation showed short erratic changes in F_0 estimates, particularly at boundaries between semitones and with added noise. An RMS moving-average filter with a 125 ms window was added to the signal processing chain to address this issue. The signal-processing strategy's output in response to harmonic complexes (detailed in the following Section 5.1.4) in the presence of white noise at increasing SNRs is illustrated in Figure 5.4.

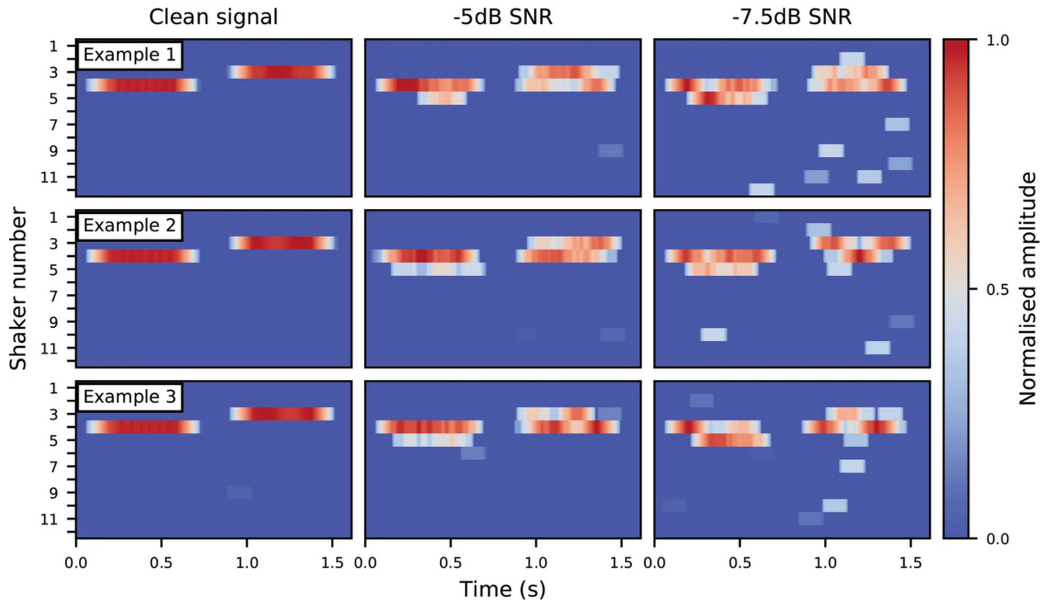

Figure 5.4: Signal-processing strategy output for harmonic complexes at 300 Hz and 300 Hz+5%. Examples 1, 2 and 3 represent 3 presentations of the same stimulus. Complexes are presented in quiet, at -5 dB SNR and at -7.5 dB SNR to demonstrate the degradation of the strategy at increasing levels of inharmonic noise[1].

In addition to the harmonic complex, sawtooth waves at 85 Hz (the lowest average F_0 of typical male speech (Hess, 2012)), 255 Hz (the highest average F_0 of typical female speech (Hess, 2012)) and 440 Hz (the standard tuning pitch for western music) were used to further assess the algorithm. A sawtooth wave was used as the prototypical waveform that has both odd and even harmonics, as is typical of many real-world stimuli such as speech. The algorithm's performance for each F_0 is illustrated in Figure 5.5.

[1]Material from: 'Fletcher et al., Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic, Scientific Reports, published 2020, Springer Nature Limited

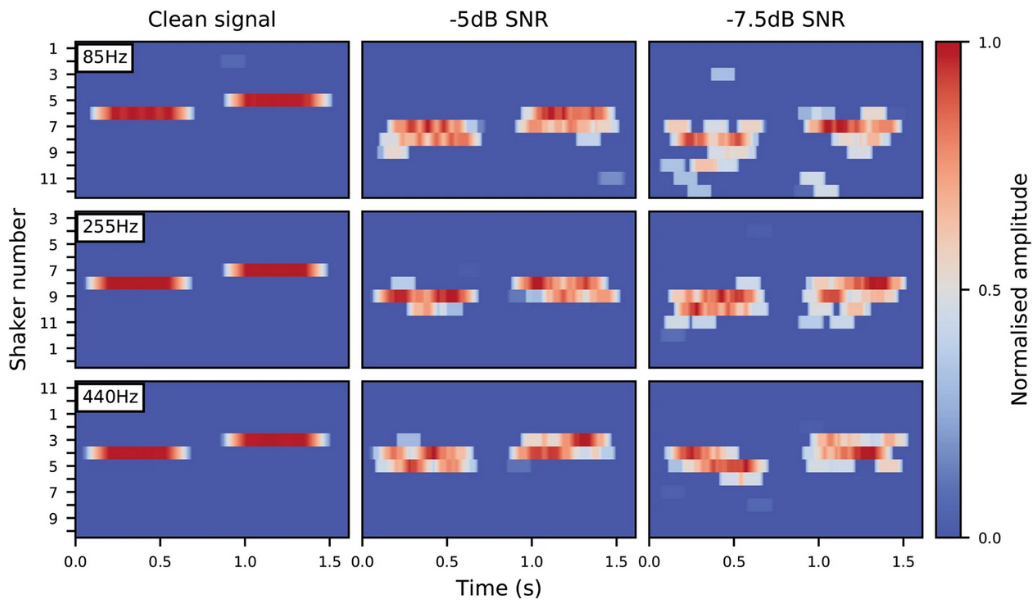

Figure 5.5: Signal-processing strategy output for sawtooth waves at 85 Hz, 255 Hz, 440 Hz and subsequently at the same frequency+5%. Sawtooth waves are presented in quiet, at -5 dB SNR and at -7.5 dB SNR to demonstrate the degredation of the strategy at increasing levels of inharmonic noise[1].

The algorithm's performance is comparable to the stimulus illustrated previously in Figure 5.4, with marginally poorer performance for the 85 Hz signal at -7.5 dB. It can also be seen that for the 85 Hz and 255 Hz signals, there is an offset of 1–2 shakers relative to the clean condition, which is due to errors produced by the initial YIN algorithm.

5.1.3 Participants

12 normal-hearing participants (3 male and 9 female, aged between 22 and 31) were recruited from student at the University of Southampton and acquaintances of the researcher. Participants were screened to ensure they:

1. Were native British English speakers.
2. Had touch perception threshold with normal limits at the fingertip (< 0.4 ms^{-2} RMS at 31.5 Hz, and < 0.7 ms^{2} RMS at 125 Hz) at the fingertip, measured using a HVLab tactile vibrometer, following ISO 13091–1:200161 specification (International Organization for Standardization, 2001). The fingertip was used as normative data is not available for the wrist or forearm.
3. Had PTA thresholds within normal limits ($<$ 20 dB HL for audiometric frequencies from 250 Hz to 8 KHz), following British Society of Audiology (2018) specification.

[1]Material from: 'Fletcher et al., Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic, Scientific Reports, published 2020, Springer Nature Limited

4. Did not present with any contraindications (as defined per British Society of Audiology (2018), as assessed using otoscopy and a health questionnaire.

Participants provided written informed consent and no payment was given for their participation. Participants reported no issues with their hearing or sense of touch, had no prior musical training and did not speak a tonal language.

5.1.4 Stimulus

An harmonic complex with an average F_0 of 300 Hz (chosen to be approximately central to the range of F_0s found in human speech (Hess, 2012) and at which pitch cues degrade for cochlear implant users (McDermott, 2004)) was used as the reference stimulus for both task familiarisation and testing sections (see procesdure Section 5.1.7 for details). The stimulus F_0 was roved by $\pm 5\%$ with a uniform distribution on each trial. The stimulus comprised of equal-amplitude harmonics, generated from the F_0 up to the Nyquist frequency (24 KHz). A 1–4 kHz 12th order (72 dB per octave) 0-phase butterworth band pass filter was then applied, removing non-pitch cues (such as brightness) that might be used to discriminate the stimuli (Mehta and Oxenham, 2017; Shackleton and Carlyon, 1994). The signal's duration was 500 ms, using 20 ms quater-sine onset and -cosine offset ramps. A 300 ms gap seperated the target and reference stimuli. The target and reference stimuli were generated using the same processing, with the exception of the target stimulus F_0 being adjusted by the adaptive track, as described in Section 5.1.7. Both stimuli were presented at a nominal level of 65 dB SPL, and roved by ± 3 dB on each presentation (with a uniform distribution) to reduce possible loudness cues. A white noise was used as the masking stimulus, to equally mask each of the harmonics of the reference and target stimuli.

5.1.5 Cochlear implant simulation.

The SPIRAL cochlear implant simulator (Grange et al., 2017) was used to simulate the pitch-perception of cochlear implant users. This simulator is a hybrid of a tone and noise based vocoder that has been shown to better match behavioural perceptual measures of cochlear implant users. The cochlear implant simulations was synthesized following procedure detailed by Grange et al. (2017), adapting the original Matlab code to provide an identical real-time Max/MSP implementation. A pre-emphasis filter was also added to this implementation, following the procedure of Fletcher et al. (2018). A schematic representation of the SPIRAL signal processing chain is provided in Figure 5.6, followed by a description of each component.

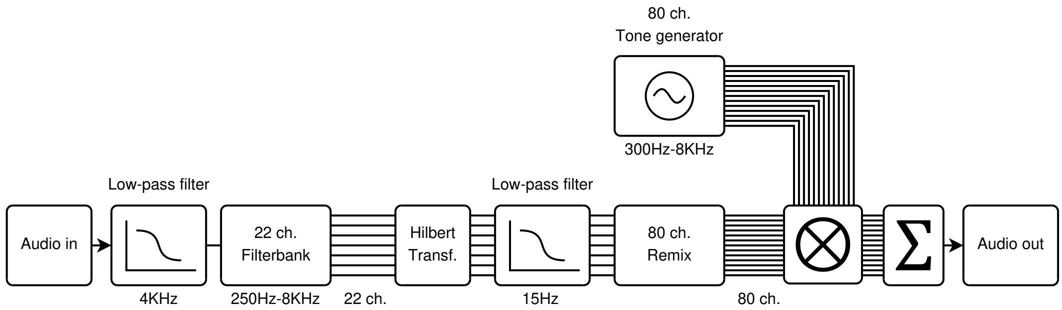

Figure 5.6: Schematic illustration of the SPIRAL cochlear implant simulation used for generating audio stimulus in real-time.

The audio stimulus is first passed through a shallow pre-emphasis filter, with a 3 dB per-octave roll off and 4 kHz cut-off frequency. This was implemented following input filter characteristics typically used in cochlear implant speech processors (Chung and McKibben, 2011). The signal was then split into 22 frequency bands using a filterbank, consisting of 22 512-point rectangular FIR bandpass filters with centre frequency spaced equally on the ERB scale. Lower and upper cut-offs for the filterbank were set at 250 Hz and 8 KHz. These filters represent the splitting of the auditory spectrum into sub-bands allocated to the 22 electrodes of a cochlear implant. The selection of 22 electrode filterbands was chosen to match those of commonly used cochlear implants from brands such as Cochlear Ltd. (Sydney, Australia). The signals from each electrode channel were then processed using a Hilbert transform and low-pass 512 point rectangular FIR filter, with a cut-off frequency of 50 Hz. This extracted the amplitude envelope, simulating the electrical stimulation provided by the cochlear implant along the basilar membrane in the cochlea. 80 tonal carriers were generated, equally spaced from 300 Hz to 8 KHz on the ERB scale. The phase of each tone was randomised. Envelopes were mixed to 80 channels, with a spread of -8dB per octave implemented to simulate the current spread of electric stimulation along the cochlea. Current spread was implemented as:

$$W(i,j) = 10^{[(S/10)\times \mathrm{abs}(\log_2(CF(i)/F(j)))]}, \tag{5.3}$$

$$M(j) = \sqrt{\sum_{i=1}^{n} \left(W(i,j) * E(i)^2\right)}, \tag{5.4}$$

where envelope i centre frequency $CF(i)$ and carrier tone frequency $F(j)$ are used to calculate a weight $W(i,j)$. n envelopes $E(i)$ are then mixed to produce the mixed envelope $M(j)$ for tonal carrier j. The mixed envelopes are then summed to produce the final output. All parameters for the vocoder were chosen to replicate those proposed by Fletcher et al. (2018).

5.1.6 Apparatus

PTA screening was performed in a sound attenuated booth conforming to British Society of Audiology standards (British Society of Audiology, 2018). PTA was performed using a Grason-Stadler GSI 61 Clinical Audiometer and Telephonics 296 D200–2 headphones. Vibro-tactile thresholds were measured using a HVLab Vibrotactile Perception Meter using a constant upward force of 2 N and a 6-mm contactor that had a rigid surround (adhering to International Organization for Standardization specifications (International Organization for Standardization, 2001)). A Brüel & Kjær (BK) calibration exciter (Type 4294) was used to calibrate the Vibrotactile perception meter. During the experiment participants responded via a iiyama ProLite T2454MSC-B1AG 24-inch touchscreen monitor. The experimenter controlled the test whilst behind a screen with no line of sight to the participant. This was to minimise the observer-expectancy effect, where an experimenter may subconsciously influence the participant's responses (Goldstein, 2011, p.374). Stimuli were generated using a custom Matlab script (The Mathwork Inc., 2019) and controlled using Max 8 (Cycling 74, 2019). A sample rate of 48 kHz was used for presentation of the audio and the haptic signals, with all signals generated by a MOTU 24Ao soundcard (MOTU, 2019). Haptic stimulation was provided by the device detailed in Section 5.1.1. Audio was presented via ER-2 insert earphones (Etymotics, 2019). A BK G4 sound level meter, with a BK 4157 occluded ear coupler were used for stimulus calibration. A BK Type 4231 sound calibrator was used for the sound level meter calibration.

5.1.7 Procedure

The experiment was completed in one session lasting around two hours. The session consisted of three phases: A screening phase, a familiarisation phase and a testing phase. The screening phase consisted of a health questionnaire to confirm participants:

1. had no conditions or injuries that may affect touch perception,
2. had not been exposed to sustained periods of intense hand or arm vibration at any time,
3. had no recent exposure to hand or arm vibration,
4. had no conditions or injuries that may affect their hearing perception,
5. had received no musical training at any time, or
6. did not speak a tonal language.

Otoscopy was performed first, to check for any contraindications that would preclude the use of insert earphones or PTA. A PTA hearing test was then conducted to ensure participants were normal hearing (hearing thresholds were < 20dB HL (British Society of Audiology, 2018)). Finally, vibrotactile detection thresholds were also measured at the fingertip, to ensure touch perception was within normal limits (see Section 5.1.3 for details).

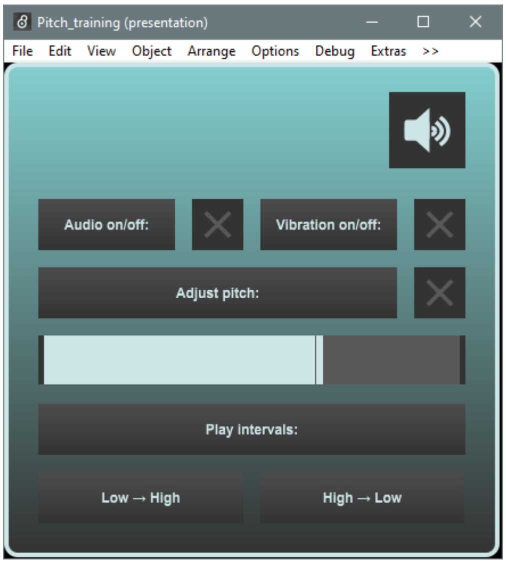

Figure 5.7: Illustration of the device familiarisation app. Audio and vibration can be toggled on and off using the first row of toggles. The pitch discrimination slider is found below and can be activated using the accompanying toggle. The interval training buttons can be found at the bottom of the app.

On passing the screening phase participants proceeded to the device familiarisation phase. In this phase, participants used an app, consisting of a pitch slider and interval training modules, to familiarise them with the haptic device. An image of the familiarisation app is provided in Figure 5.7. In each module, participants could select either haptic only, audio-haptic and audio only modes. The pitch slider module played a constant tone. Participants adjusted the frequency of the tone based on a slider position. For the interval training module, participants selected either a "Low → High" or "High → Low" button. These determined the order of pitches for two consecutive tones. Pitches were randomised, with intervals ranging from a single semitone up to 10 semitones. Participants could repeat presentation as many times as they wished, but stimuli intervals were randomised on each repetition. In both modules, stimuli were presented clean, with no cochlear-implant simulation applied. Participants were familiarised for around 5–10 minutes, and were able to ask the experimenter questions for clarification.

Having been familiarised with the device, participants were then familiarised with the task to be used in the testing phase. For each condition participants completed 15 trials of the task. A two-alternative forced-choice task was used for task familiarisatoin and testing. Participants were required to identify which interval had the stimulus with the higher pitch using the on-screen interface illustrated in Figure 5.8.

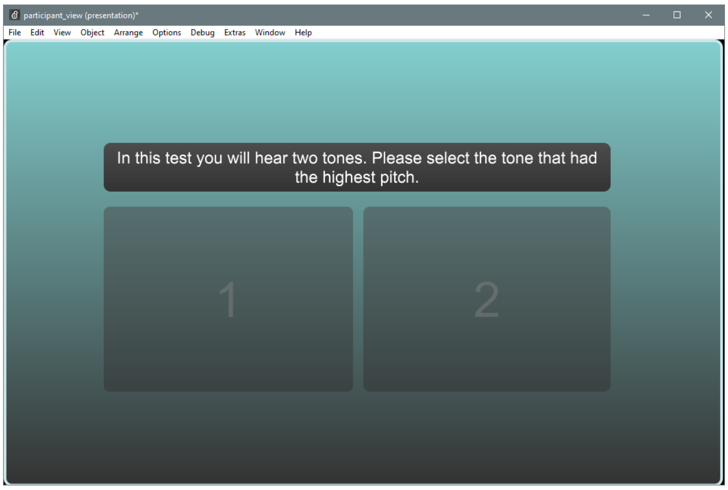

Figure 5.8: Illustration of the display used for the participants to provide feedback during testing. Instructions read "In this test you will hear two tones. Please select the tone that had the highest pitch."

The panels "1" and "2" signified if the first or second stimulus was higher in pitch. On selecting either panel, the next trial was presented with visual feedback provided to the participant indicating whether their response was correct. A one up, two down adaptive track varied the difference in pitch between the target and reference stimulus. The pitch difference between the reference and target stimulus decreased on two consecutive correct responses and increased after every incorrect response. An initial difference of 80% was set, with track steps of 10% for the first 2 trials, 5% for the third and 1% for the final four trials. A threshold was calculated as the average of the final 4 reversals. There were three stimulus conditions: Audio only, combined audio-haptic and haptic only. These conditions were counterbalanced across participants to account for any effects of condition order. Additionally there were three noise conditions: clean signal, white noise at -5 dB SNR and white noise at -7.5 dB SNR. These were presented in a random order for each condition. All parameters and condition selection was performed automatically using a custom Max patch (illustrated in Figure 5.6) to minimise opportunity for errors by the experimenter.

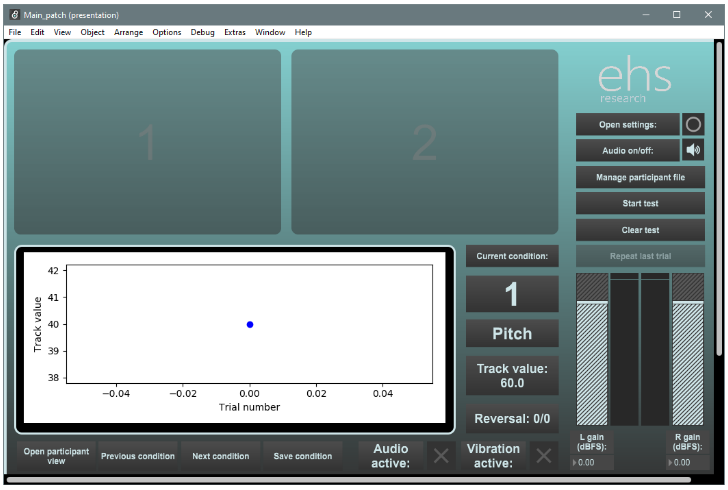

Figure 5.9: Illustration of the display used by the experimenter for test control and monitoring. Top panels indicate responses from the participant. The middle panel displays the adaptive track for each condition as the test progressed. Controls at the bottom of the app allow for wiching of conditions, with current number, track value and number of completed reversals displayed alongside.

The experimental protocol detailed was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 47769). All research was performed in accordance with the relevant guidelines and regulations.

5.2 Statistics

Data was not normally distributed (as assessed using a Shapiro-Wilks test). As a result, data was analysed using non-parametric statistical tests. Two Friedman tests (applying Bonferoni-Holm correction for multiple comparisons) were used to assess overall condition effects. Data were further analysed using three planned post-hoc Wilcoxon signed-rank tests (again applying Bonferoni-Holm correction) for comparison of individual conditions.

5.3 Results

Pitch discrimination thresholds for each condition (audio only, audio-haptic and haptic only) are illustrated in Figure 5.10. Each condition displays threshold for the 3 noise conditions (clean signal, -5 dB and -7.5 dB SNR white noise).

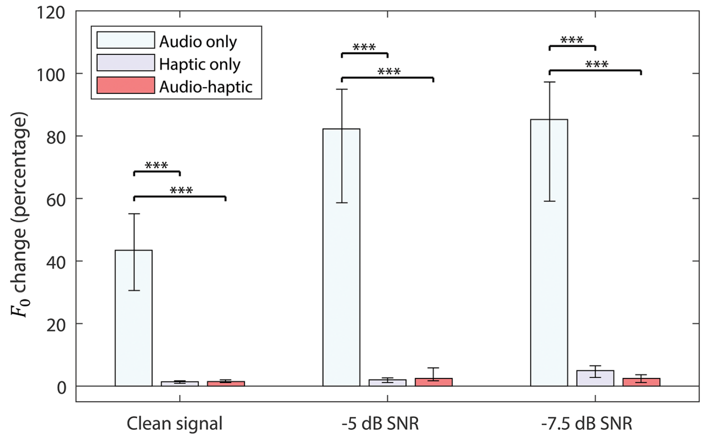

Figure 5.10: Pitch discrimination thresholds for the audio only, audio-haptic and haptic only conditions for the 12 participants. Conditions with no background noise and with background noise at either -5 dB or -7.5 dB signal-to-noise ratio (SNR) are shown. Bars represent the median F_0 threshold. Error bars show 5% and 95% confidence intervals (bootstrapped for each condition using 1000 samples with replacement)

Friedman tests for repeated measures were conducted with factors: stimulation type (audio only, haptic only, audio-haptic) and noise type (clean, -5 dB SNR, and -7.5 dB SNR). A significant overall effect of stimulation type was found ($\chi^2(2) = 18.17, p < 0.001$). Additionally, a significant overall effect of noise was found for both audio only ($\chi^2(2) = 18.17, p < 0.001$) and haptic stimulation only ($\chi^2(2) = 15.45, p = 0.001$). For the audio only condition the median threshold was 43.4% without noise (ranging from 8.4% to 106.0% across participants), increasing to 82.2% with noise at −5 dB SNR (ranging from 27.6% to 130%) and to 85.2% with noise at −7.5 dB SNR (ranging from 29.7% to 116.5%). For the haptic only condition the median threshold was 1.4% without noise (ranging from 0.8% to 3.5%), 2.0% in noise at −5 dB SNR (ranging from 0.6% to 6.6%), and 5.0% in noise at -7.5 dB SNR (ranging from 1.1% to 10.8%). There was no significant overall effect for audio-haptic stimulation ($\chi^2(2) = 2.09, p = 0.35$). For the audio-haptic condition, the median threshold was 1.5% without noise (ranging from 0.8% to 4.1%), 2.5% with noise at −5 dB SNR (ranging from 0.8% to 5.5%), and 2.4% with noise at −7.5 dB SNR (ranging from 0.9% to 15.0%).

Post-hoc Wilcoxon signed-rank tests (with Bonferroni-Holm correction applied for multiple comparisons) showed that pitch discrimination improved significantly with audio-haptic stimulation than with audio alone ($T = 78, p = .001, d = 3.76$). Threshold improvments of 42.0% were found without noise (ranging from 7.5% to 103.4% across participants), of 80.2% with noise at −5 dB SNR (ranging from 7.5% to 103.6%), and of 80.3% with noise at −7.5 dB SNR (ranging from 5.7% to 95.2%). Pitch discrimination thresholds also improved significently for the haptic alone condition, when compared to the audio alone condition ($T = 78, p = .001, d = 3.75$). The median threshold was 41.9% better without noise (ranging from 7.2% to 104.9% across participants), 79.8% better in noise at -5 dB SNR (ranging from 7.0% to 101.0%) and 80.8% better in noise at -7.5

dB SNR (ranging from 7.1% to 91.0%). Pitch discrimination between haptic alone and audio-haptic stimulation showed no significant difference ($T = 35, p = .791, d = 0.05$)

5.4 Discussion

The results show that use of the device improved pitch discrimination performance considerably for simulated cochlear implant users. Without noise, participants achieved median F_0 discrimination thresholds of 1.4% without noise. This is markedly less than the semi-tone minimum target that was expected when designing the device. Results showed that even the worst performers achieved discrimination thresholds less than the this target, achieving similar pitch discrimination performance to that of the best performing cochlear implant users (Drennan et al., 2015; Kang et al., 2009). For the audio haptic and haptic alone conditions, the highest performing participant were able to discriminate differences of as little as 0.8%. This is comparable to normal hearing listeners assessed with similar stimulus (however pitch discrimination threshold are highly sensitive to the precise properties of the stimulus (Kaernbach and Bering, 2001). The reported improvements in pitch discrimination suggest that the device could be used to provide segmental and suprasegmental, as well a prosodic, cues, which cochlear implant users currently have limited access to. Through provision of these additional important features of speech, this method could aid speech-in-noise performance for cochlear-implant users.

The notable noise robustness of the method further suggests that this type of pitch mapping may be viable for real-world applications. The lack of an effect of noise for the audio-haptic condition, and the low median thresholds for the haptic alone conditions (rising from 1.4% without noise to only 5.0% at -7.5 dB SNR) highlight the effectiveness of the signal processing applied to handle background noise. The performance of the proposed strategy is particularly notable given that traditional real-time pitch extraction approaches are typically susceptible to background noise (Jouvet and Laprie, 2017). Results also suggest that this approach is usable in far less favourable conditions than a cochlear implant would be usable in, as at -7.5 dB SNR cochlear-implant users are unable to perform pitch discrimination tasks (Kreft et al., 2013) or speech recognition tasks (Wilson, 2015; Fletcher et al., 2019; Huang et al., 2017). However, it should be noted that this approach has only been assessed in inharmonic background noise. This suggests the approach would be robust to real-world noises such as wind or rain, but the effects of noise with competing harmonic elements such as competing speakers should also be evaluated.

Additionally, the underlying mechanisms of the observed sub-semitone performance is not currently clear. Figures 5.4 and 5.5 show that at increasing SNRs the pitch estimates fluctuate around the true frequency channel. Participants could therefore be comparing distributions of activated shakers in order to discriminate the stimuli. It is thought that a similar process contributes to signal detection in the auditory system (Verhey et al., 2007; McDermott et al., 2013).

The absence of a degradation in performance between the audio-haptic and the haptic-alone conditions suggests that the substantially poorer and more distorted pitch cues provided by the audio did not distract the participants. There was no further improvements seen in the audio haptic condition than the haptic alone condition. This is expected given the inferior cues provided by the audio, as illustrated in the poorer performance of participants in the audio-alone condition. As described in Section 2.1.1.5, the principal of inverse effectiveness states that the highest levels of multisensory integration occur when individual modalities provide low-quality information — a requirement that has not been met here.

Their are also a number of limitations to this study that should be considered. Most notably that this study did not asses the benefit of providing pitch cues via haptics for speech-in-noise performance. However, previous findings (such as Huang et al. (2017) and Fletcher et al. (2018), detailed in Section 2.4) suggest that providing pitch and spectral based cues can result in considerable improvements to cochlear-implant user's speech-in-noise performance. Further work in this area is also detailed in Chapter 6.

A further consideration is that the results presented do not demonstrate an improvement in auditory pitch perception, but show that it is possible to provide detailed pitch cues that can be detected via haptic stimulation on the forearm. Further work is required to determine the degree to which integration with auditory stimulus occurs for the type of tactile pitch cues described in this chapter, and the degree to which better integration may improve performance.

Limits were also placed on the stimulus, as this was restricted to a single reference F_0 of around 300 Hz (roved by $\pm 5\%$). Section 5.1.2 details the analysis of outputs for different reference F_0s, suggesting that the method was robust to change in F_0s and therefore that results would likely be generalisable to a range of F_0s. A further limitation is that this study was performed on normal-hearing participants, who listened using a cochlear implant simulation. Whilst every effort was made to match the performance of the cochlear implant simulation accurately (See Section 5.1.5), perception of stimulus is likely to vary between cohorts. Further work is needed to understand the effects of this stimulation in actual cochlear-implant users.

Finally, a lack of training may also limit the benefits presented. Previous studies have shown that training can improve performance for auditory frequency discrimination tasks, when two hour of training is given per day (Moore, 1973). In addition haptic enhancement of speech-in-noise performance has also been shown to benefit from training regimes (Fletcher et al., 2019; Fletcher et al., 2018). Therefore the current results may underestimate the absolute limits of pitch perception using this method and further work is needed to determine the potential benefits of training on these results.

Further work is needed to adapt the presented methodology in the development of a real-world device for improving speech-in-noise performance. It has been argued that the two-point discrimination thresholds (used to select the 3 cm spacing between motors) may overestimate the minimum spacing for separable vibration sources (Bach-y-Rita, 2004). Future designs may allow for closer spacing of motors, which could allow for a more compact device that does not require use of the full forearm. In addition, the current signal processing strategy provides only frequency cues, modulated by the broadband envelope. By combining these with transmission of other cues such as narrowband envelopes (as demonstrated to be effective by e.g. Fletcher et al. (2018)) may offer additional speech-in-noise benefits to that found by Fletcher et al.

The results of this study have demonstrated the efficacy of the device presented to provide fine-grained pitch information that would otherwise not be accessible via a cochlear implant. These results have been shown to be robust to high levels of non-harmonic background noise, representative of many real-world competing noises. Stimulation was provided on a site feasible for real-world implementation of this style of device. The results presented suggest that this type of device could be used to provide a non-invasive and inexpensive means to improve speech-in-noise performance for cochlear-implant users.

5.5 Contributions

The data used in this study was originally collected by Nour Thini in completion of her MSc in Audiology at The University of Southampton. All device development, experiment design, stimulus processing and data analysis were carried out by the author Samuel Perry and supervisor Mark Fletcher. Statistical analysis was performed independently of the original analysis for the MSc. This chapter was published in the peer-reviewed article Fletcher et al., 2020c.

Chapter 6

Improvement of speech-in-noise performance for cochlear-implant users using electro-haptic stimulation

Section 2.3 highlighted the limitations of cochlear implants to effectively transmit intelligible speech in even small levels of background noise. This section also suggested that the use of haptic-based interventions may provide additional benefit to cochlear implant users, with previous studies showing considerable benefit of applying haptic stimulation for clinical scores of speech-in-noise performance. In particular, studies have shown improvements of 13.9% words correct scores in colocated noise using a vocoder based approach (Fletcher et al., 2019), 2.2 dB SNR improvement for co-located noise using an F_0 extraction approach (Huang et al., 2017) or 2.8 dB SNR improvement for spatially-separated noise (Fletcher et al., 2020b). However, previous studies have been limited in their implementations. In studies such as those by Fletcher et al. (2020b) and Fletcher et al. (2019), the devices used to deliver the haptic feedback were lab-based devices that are unsuitable for use in the real-world. The study by Huang et al. (2017) used haptic-aids that were mounted on the fingertip (limiting their potential for real-world application, due to the cumbersome properties of finger-mounted devices). In addition the signal processing strategy only provided F_0 cues, removing important cues for speech intelligibility such as information on the spectral shape of the speech (including higher formants and the broader spectral envelope morphology) (Hillenbrand et al., 2006).

This Chapter aims to build on the results presented in previous studies by developing and assessing a new device and accompanying signal processing strategy for the delivery of haptic stimulation. The device is designed to address a number of the outlined flaws in previous methodologies, to provide a readily translatable method for improving cochlear implant user's speech-in-noise performance. Sections 6.1.1 to 6.1.3 outline the development of the mosaicOne_C device and signal processing strategies designed to improve speech in noise performance. Sections 6.1.4 to 6.1.7 outline an experiment to assess the benefits of the device for speech-in-noise performance in simulated cochlear implant users (see 5.1.5 for implementation of the cochlear implant simulator). In addition the proposed modifications to the signal processing strategy are evaluated against an adaptation of the traditional vocoding method presented by Fletcher et al. (2018). This study aims to answer the following research questions:

1. Does providing haptic stimulation using frequency-focused vocoding improve speech-in-noise performance more than the traditional approach?
2. Does extensive training increase the benefit of haptic enhancement for either the original vocoder or frequency focused approach?

The results of this study are detailed in Sections 6.2 and 6.3. The potential for the proposed device and signal processing strategy and current limitations of the method as an intervention for cochlear implant users are outlined in Section 6.4.

6.1 Methods

6.1.1 Tactile stimulation device: The mosaicOne_C

A 4 motor haptic stimulation device (the mosaicOne_C, illustrated in Figure 6.1) was developed to produce precise haptic stimulation to participants using the signal processing strategies detailed in Sections 6.1.2 and 6.1.3.

Figure 6.1: Image of the custom made mosaicOne_C haptic device. The device houses 4 306–10H 7–mm haptic motors, spaced equally around the wrist.

The device consisted of 4 Precision Microdrives 306-10H 7-mm vibration motors, spaced equally around the wrist. Each motor was connected to a Texas Instruments DRV2605L haptic driver, used to convert control voltages to current for driving the motors. The DRV2605L was chosen for its flexibility. The driver is capable of automatic calibration and motor control routines, which were set to maximise the performance of the haptic motors. The driver was also suitable for rapid prototyping given the ability to simply program the device using an Arduino via I2C. Table 6.1 details the parameters used to initialise all haptic drivers.

Parameter	Setting	Notes
Input mode	Analog input mode	Allow for control of motors directly using control voltages from a DC coupled soundcard
Waveform playback mode	Closed-loop unidirectional ERM mode	Maximise driver performance using 0–1.8V range to drive motors, applying overdrive and active braking based on auto-calibration results
Calibration time	1.0-1.2 seconds	Maximum allowable by the device
Startup boost	Enabled (default)	Extra gain applied for transient signals to improve response of motors
Noise gate	Disabled (0%)	Remove signal distortion caused by gating of low level voltages (noise reduction applied in signal processing)
Rated motor voltage	3.0V	As per 305–10H datasheet
Maximum operating voltage	3.6V	As per 305–10H datasheet
Drive time	4.8ms (208.33 Hz; default)	Sample-rate for back-EMF detection
Back-EMF gain	1.8× (default)	
Brake factor	4× (default)	
Brake stabiliser	Enabled (default)	

Table 6.1: DRV2605L haptic driver parameters

All haptic driver parameters were controlled using an Arduino Uno via I2C. An I2C multiplexer (Texas Instruments TCA9548A) selected each driver sequentially, applying parameters and running each auto-calibration routine in turn. Four control voltages were generated using a DC-coupled MOTU 24Ao soundcard. Each control voltage drove the analog input pin of one motor – channels were mapped from the lower left of the wrist (channel 1), clockwise around the participant's wrist, to the lower right (channel 4). The complete experiment setup, including audio stimulus hardware, is illustrated in Figure 6.2

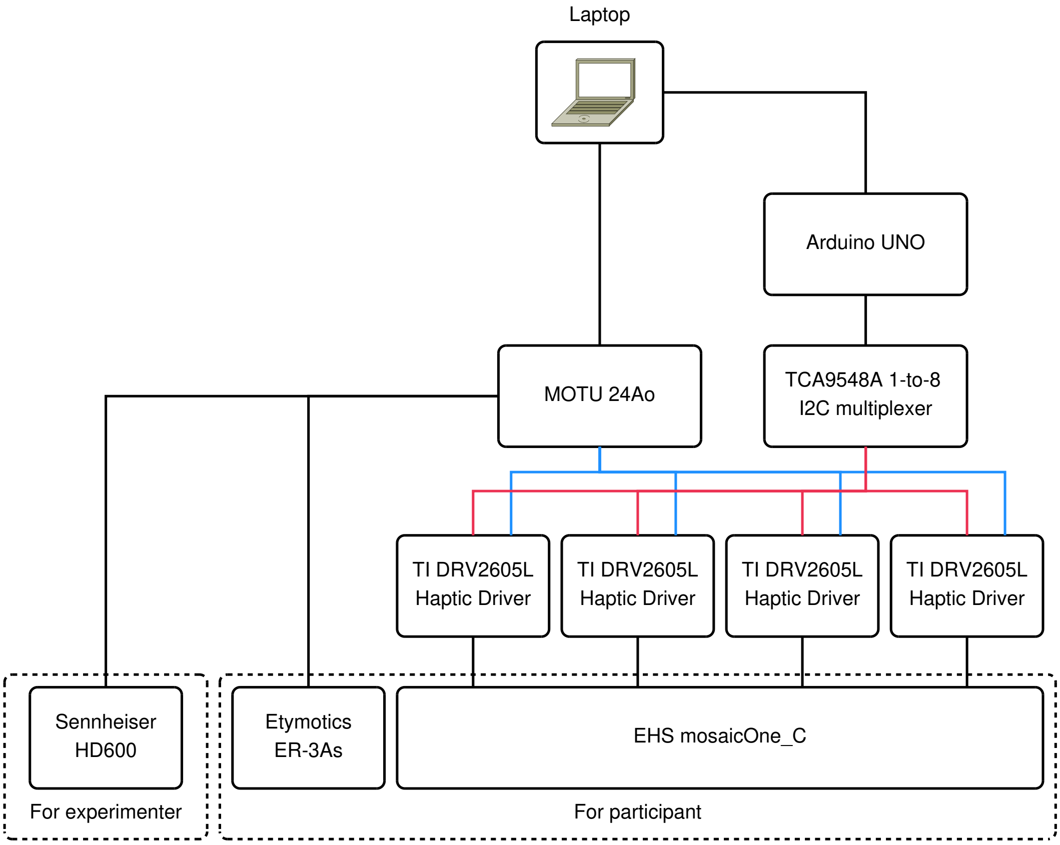

Figure 6.2: Schematic illustration of the experiment setup. Auditory and tactile signal processing is performed by the MOTU 24Ao. An arduino and I2C multiplexer are connected for initialisation and calibration of haptic motor drivers. Etymotic ER-3As provide auditory stimulus to participant and Sennheiser HD600s provide masking noise for the experimenter.

6.1.2 Tactile signal-processing: Original vocoder.

Two signal processing strategies are proposed to convert audio to haptic stimulation. The first is an adaption of a DSP strategy originally proposed by Fletcher et al. (2018). Fletcher et al.'s vocoder strategy has been shown to effectively provide stimulation for the improvment of speech-in-noise for cochlear implant users. However, the approach uses a single contactor and seperates audio bands across carrier frequencies, as opposed to across motors spatially. Previous Chapter 5 suggested that haptic stimulation could deliver perceptual cues effectively using a spatial mapping. This allows for design of mapping with a resolution limited primarily by the design space of the user's wrist. Given the limitations of inexpesive haptic motors (that cannot provide independent modulation of frequencies, as described in Fletcher et al. (2018)), this may be a more suitable alternative that could be applied in development of a real-world device. Therefore, this approach has been adapted to provide a spatial mapping of the extracted sub-bands. First, audio is downsampled to 16 kHz and split into 4 sub-bands using 4 512-point rectangular FIR bandpass filters, equally spaced on the ERB scale, with a lower cutoff of 100 Hz and upper cutoff of 1 kHz. Envelopes for each band were then extracted using a absolute Hilbert transform and a 512-point rectangular FIR lowpass filter, with a cutoff frequency of 15 Hz, as set for the original vocoder. A multi-band dynamic-range expander was also implemented as a simple noise reduction stage, adapting the approach of Fletcher et al. (2018) to be implementable in real-time. The expander threshold was set at -0.1 dB below the envelope's average amplitude over the previous 500 ms. The expansion ratio was set at 9.0, the attack was set at 4 ms and the release was set at 10 ms. No lookahead was used to maximise synchronisation between the haptics and the audio. The resulting signals were output to each of the 4 motors. A schematic illustration of the signal processing is provided in Figure 6.3.

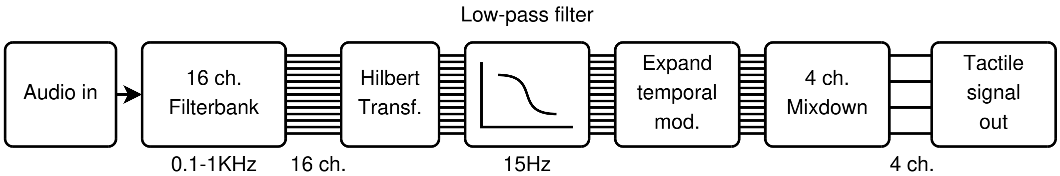

Figure 6.3: Schematic illustration of the vocoder approach, adapted from Fletcher et al. (2018), used for generating tactile stimulus from input audio in real-time.

6.1.3 Tactile signal-processing: Frequency-focused vocoder.

This signal processing strategy was developed as a further adaption of the vocoder implementation detailed in Section 6.1.2. The original vocoder filterbank channels were too broad to present find-grain pitch information for voiced speech. As pitch is such a key feature of speech perception (see Chapter 2.1 for details), a vocoder that is able to dynamically focus on the fundamental frequency of voiced speech may provide additional benefits for speech-in-noise performance.

To create the haptic signal, first audio is downsampled to 16 kHz. The signal is then processed by a filterbank, and separately in parallel by a real-time implementation of the YIN fundamental frequency estimation algorithm. YIN is implemented following the procedure detailed by de Cheveigne and Kawahara (2002). F_0 and harmonicity estimates were produced using a window size of 1024 samples, with a hop size of 256 samples. F_0 values with a harmonicity less than 0.1 were removed (unvoiced or silent segments). A running mean of the last 90 harmonic estimates is then calculated, producing a smoothed F_0 contour that is used to select the filterbank sub-bands. The center sub-band is selected by taking the sub-band index i_3 where $SF_{lower}(i_3) \leq F_0 > SF_{upper}(i_3)$, defining $SF_{lower}(i)$ as the lower cutoff frequencies of the filterbank and SF_{upper} as the upper cutoff frequencies at index i. Adjacent lower and upper bands were then selected as $i_1 = i_3 - 2, i_2 = i_3 - 1, i_4 = i_3 + 1, i_5 = i_3 + 2$. As with the "original" vocoder, envelopes for each band were then extracted using an absolute Hilbert transform and a 512-point rectangular FIR lowpass filter, with a cutoff frequency of 15 Hz. A temporal expander is again implemented, using a threshold set at -0.1 dB below the envelope's average amplitude over the last 500 ms. The expansion ratio is set at 9.0, the attack is set at 4 ms and the release is set at 10 ms. No lookahead is used. A schematic illustration of the signal processing approach is provided in Figure 6.4

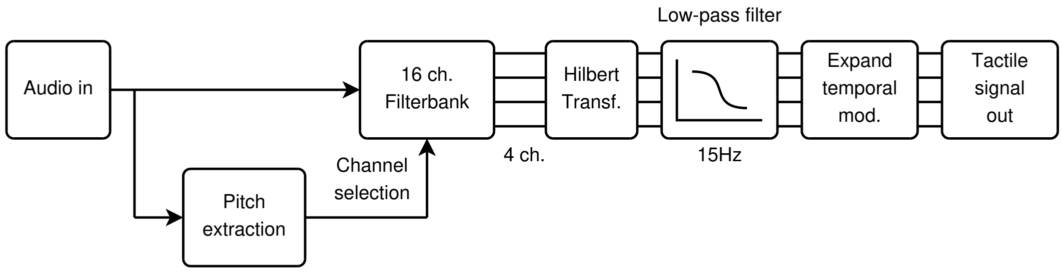

Figure 6.4: Schematic illustration of the "frequency-focused" vocoder approach used for generating tactile stimulus from input audio in real-time.

6.1.4 Participants.

6 normal-hearing participants (4 male and 2 female, aged between 20 and 35) were recruited from student at the University of Southampton and acquaintances of the researcher. Participants were screened to ensure they:

1. were native British English speakers
2. had touch perception threshold with normal limits ($< 0.4\text{ms}^{-2}$ RMS at 31.5Hz, and $< 0.7\text{ms}^{2}$ RMS at 125Hz) at the fingertip
3. had PTA thresholds within normal limits (< 20dB HL for audiometric frequencies from 250 Hz to 8 kHz), following (Insert BSA standard) specification
4. did not present with any contraindications, as assessed using otoscopy and a health questionnaire

Participants that met these criteria were then familiarised with the CI simulator. 10 male speaker sentences from lists 13 and 14 of the IHR familiarisation corpus (detailed in Section 6.1.5) were presented using the cochlear implant simulation, controlled by the participant using a touchscreen interface. Sentences were displayed alongside the audio files, as illustrated in Figure 6.5. Participants were asked to listen to each sentence, and were able to repeat sentences as many times as they wished.

Figure 6.5: The cochlear implant simulation familiarisation interface

Following the CI simulator familiarisation, participants were presented with up to 5 BKB lists in quiet via the CI simulator. Participants that could not achieve 70% words correct score after 5 lists were excluded from further testing. No participant was exclused as a result of this criteria.

6.1.5 Stimuli.

This study used two separate speech corpora: Screening and test sessions used BKB Male, Female and Institute for Hearing Research BKB familiarisation sentences (male talker). This provided a total of 60 male and 33 female lists of sentences, 10 sentences per list. The 2 lists that included sentences used for familiarisation were excluded for all further testing. Training sessions used the ARU IEEE sentence lists, which comprises of 72 lists of 10 sentences, spoken in a variety of British male and female accents. From this corpus, the male Kent, male Avon, female Berkshire and female Middlesex lists were selected to provide two accents for male and female training. Multiple talkers were used to train participants across a variety of speech styles (with varying acoustic and linguistic properties such as pitch, tone and inflection for example). Selection of lists from each corpus is detailed in Section 6.1.7. All speech stimuli were presented via a real-time implementation of the SPIRAL vocoder (detailed in Section5.1.5) at 60 dB LAeq.

The SPIRAL cochlear implant simulator was used to simulate the speech-perception of CI users. This vocoder aims to provide an accurate simulation of the effects of current spread in the cochlear. SRTs for normal-hearing listeners using the SPIRAL vocoder have been shown to better match those of CI users than for traditional vocoder approaches. The implementation used matches that used in previous chapters — details on the implementation can be found in Section 5.1.5.

6.1.6 Apparatus.

The experiment was conducted in a quiet listening room. All auditory stimulus was presented to the participant using Etymotics ER-2s, via a MOTU 24Ao soundcard. Stimulus was calibrated using a B&K G4 Type 2250 sound level meter (Brüel & Kjær, 2021a), with a B&K 4157 occluded ear coupler (Brüel & Kjær, 2021b). Sound level calibration was performed using a B&K Type 4231 sound calibrator (Brüel & Kjær, 2021c). The experimenter was also presented masking noise via Sennheiser HD600 Headphones, via the MOTU 24Ao in order to reduce the observer expectancy effect (as mentioned in Chapter 5). Tactile detection thresholds were measured using HVLab diagnostic software and a HVLabs tactile vibrometer. The vibrometer had a 6mm probe and a rigid surround, following ISO specification for measurement of tactile thresholds (International Organization for Standardization, 2001). All other haptic stimulus was presented using a mosaicOne_C haptic stimulation device (as described in Section 6.1.1) fitted to the wrist of the participant's dominant hand. The mosaicOn_C was calibrated at the start of each session, calibrating each motor sequentially as detailed in Section 6.1.1.

6.1.7 Procedure.

A schematic illustration of the procedure is provided in Figure 6.6 and participant screening criteria is provided in Section 6.1.4. Participants that passed the screening and CI simulator familiarisation sessions continued to subsequent CI simulator SRT measurement, device calibration, condition familiarisation, testing and training sessions.

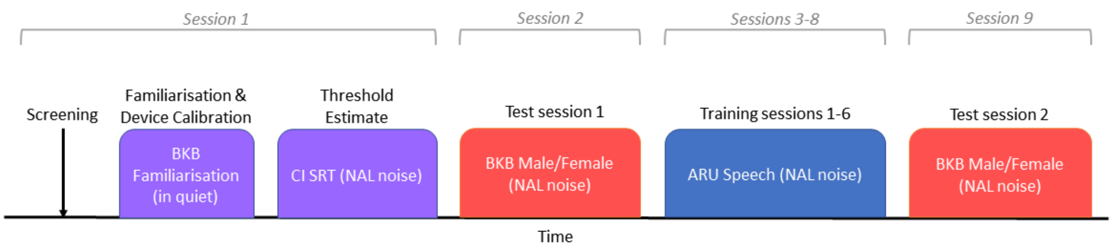

Figure 6.6: Schematic illustration (not to scale) of the experiment timeline.

mosaicOne_C fitted to participant and calibrated using procedure detailed in Section 6.1.1. To account for variability in wrist sensitivity between participants, a short amplitude matching routine was performed at the start of session 1. Participants were presented with a screen and instructions as illustrated in Figure 6.7.

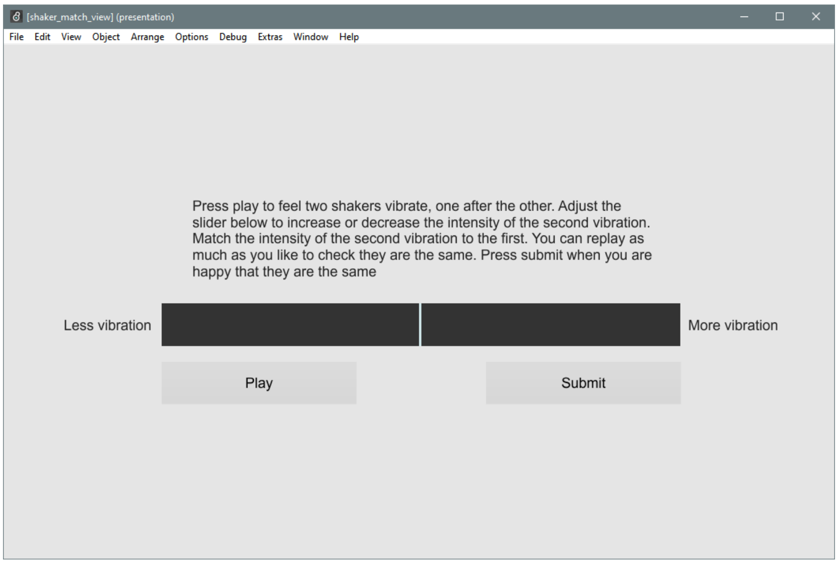

Figure 6.7: Participant view of the motor calibration procedure. Interaction with the interface was performed using a touchscreen.

During the calibration procedure participants were instructed that they would feel two sequential vibrations when they pressed the “play” button. The task was to set the slider so that both vibrations were the same intensity. Participants could repeat the vibration sequence as much as they wished, but needed to adjust the slider and play the stimulus at least once before submitting a response. On each play, a motor was ramped to a comfortable level (determined during piloting) using a raised cosine ramp of 20 ms, sustained for 560 ms and ramped off using a raised cosine ramp for 20 ms, giving a total stimulus duration of 600 ms. A second motor then presented the same stimulus initially 3 dB higher than the comfortable level or lower than the first. Absolute intensity of both stimuli were varied randomly with a uniform distribution by ±4dB on each playback to avoid use of absolute intensity cues for completion of the task. Every combination of motors was presented twice. Once with the first motor starting 3 dB higher, and once with the first motor starting 3 dB lower. Presentation order was randomised across participants. Presentations for each motor combination were then averaged to provide a map of perceived relative amplitude differences between all motors. Offsets were applied based on these differences to create perceptually uniform stimulation across motors for the remainder of the experiment.

The experiment then had an initial session to familiarise participants with each of the three conditions: Audio only, frequency focused vocoder and original vocoder. The order of conditions were fully counterbalanced across participants. Participants were told that there were two vibration conditions (labelled “Vibration 1” and “Vibration 2”) and one audio only condition. For the vibration conditions, the condition labels were alternated between participants to account for any biasing towards a particular label. Additionally, the experimenter was blinded to the current condition throughout the process to avoid any unintended biasing. Participants were presented with the same 10 IEEE sentences as for CI familiarisation. These sentences were excluded from any future training sessions. Participants were allowed as many repetitions as requested, continuing to the next condition at their request. No participant spent more than 5 minutes per condition.

On completion of condition familiarisation, participants proceeded to the CI SRT estimation session. In this session participants were played sentences mixed with background noise via the cochlear implant simulator and instructed to repeat what was said. A single adaptive track was used with a 1 up, 1 down procedure. The track used a step size of 5dB for the big reversal, 2.5dB for the medium reversals and 1dB for the small reversals. There was a total of 1 big reversals, 2 medium reversals and 6 small reversals used. Speech stimuli was set to a nominal level of 60 dB LAeq. In the first trial, noise stimulus was presented at +20 dB SNR and was then adjusted based on the adaptive track after each response. BKB Male lists 1–10 were used for speech stimulus. Piloting indicated that this would be sufficient for a complete track and no participant exceeded this number of lists. A multi-talker party noise recorded by the National Acoustic Laboratories (Keidser et al., 2002) was used. The noise was recorded at a real-world party and had been filtered to match the international long term average spech spectrum (Byrne et al., 1994). The amplitude of the noise was varied based on the adaptive track. The onset of the presented sentence was varied randomly by 200–400ms after the onset of the noise stimulus. This was to ensure participants could not take advantage of timing cues when completing a trial. On completion of the track, the SRT was calculated as the average of the final 6 reversals. Instructions were provided verbally and additionally via a screen which instructed participants when to listen and when to respond. No feedback was provided to indicate correctness of the participant's response for this session.

The testing/training section consisted of 2 test sessions and 6 training sessions, each lasting around 45 minutes. A gap of no more than 72 hours was permitted between any two sessions. In the pre- and post-training test sessions speech-in-noise performance was measured using 6 BKB male and 6 BKB female lists per condition. List presentation was counterbalanced, with either all female or all male lists presented to the participant first. This was to avoid order effects across participants based on the speaker type. As with previous sessions, participants were played sentences mixed with background noise via the cochlear implant simulator and instructed to repeat what was said. However, for testing NAL noise was set at a constant level for each session. Noise was presented at an SNR equivalent to the participants 50% SRT. This performance matching per participant aimed to avoid ceiling or floor effects in responses. Instructions were provided verbally and additionally via a screen which instructed participants when to listen and when to respond. No feedback was provided to the participants to indicate correctness of their response. Onset of sentence in noise was varied randomly by 200–400ms to ensure participants could not use any timing cues. White noise was presented to the experimented at a comfortable level of around 65 dBA during trial presentations, in order to mask any auditory cues that may indicate the current condition. The participant's performance was measured as percent words correct.

For the 6 training sessions, participants were presented with IEEE sentence lists 1–72 at set noise levels. 36 lists were spoken by one of the two female speakers and the other 36 by one of the two male speakers. In total, 20 female Berkshire, 16 female Middlesex, 20 male Kent and 16 male Avon sentences were used. These lists were divided between the 3 conditions, providing a group of 24 lists per condition. These groups were created by selecting 4 lists from a single male speaker and an additional four from a single female speaker at a time, sampling without replacement. Choice of speaker within a gender was selected at random. As a result, an equal number of male and female lists were presented per condition across all sessions. Each session used four lists per condition. For each condition, the four lists were spoken by the same speaker to maintain continuity of content. The order of speakers for each condition was also randomised across sessions. NAL noise was set at a constant level for each session. It was set to the participant's SRT+5dB for sessions 1 and 2, to their SRT+2.5dB for the second 2 sessions and to their SRT for the final 2 sessions, following the procedure outlined in Fletcher et al. (2019). These levels were selected both to match performance per participant and to gradually increase the difficulty of session, allowing participants to habituate to each condition. As with training further stimulus presentation followed that of the training sessions. Additionally participants were provided on screen feedback after every response, indicating the correct sentence.

The experimental protocol detailed was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (ERGO ID: 53690.A1). All research was performed in accordance with the relevant guidelines and regulations.

6.2 Statistics

Shapiro-wilks test confirmed normality of data for the overall (Audio only: $W = 0.97, p = 0.07$; Frequency-focused: $W = 0.99, p = 0.45$; Original vocoder: $W = 0.98, p = 0.25$), female only (Audio only: $W = 0.98, p = 0.68$; Frequency-focused: $W = 0.96, p = 0.30$; Original vocoder: $W = 0.97, p = 0.42$) and male only (Audio only: $W = 0.96, p = 0.10$; Frequency-focused: $W = 0.97, p = 0.22$; Original vocoder: $W = 0.98, p = 0.43$) analyses. Maulchy's test indicated that sphericity had not been violated for any factor, therefore no correction for this was applied. A 2-way repeated measures ANOVA was performed with factors "Condition" (Audio only, Frequency focused and Original vocoder) and

"Session" (before and after training) to determine if an effect was present between the factors.

2 planned post-hoc 2-way repeated ANOVAs (with Bonferonni-Holm correction for multiple comparisons) were also conducted to explore the effect of Condition and Session for speaker genders separately. Mauchly's test indicated that the assumption of sphericity had been violated for the female speaker only condition factor ($\chi^2(2) = 9.11, p = 0.01$), so degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity for this factor. t-tests were then performed for the female speaker only data, correcting for multiple comparisons using Bonferonni-Holm correction. Further unplanned t-tests were performed to explore the relationship between each condition on an individual level. No correction for multiple comparisons was applied to these exploratory tests.

6.3 Results

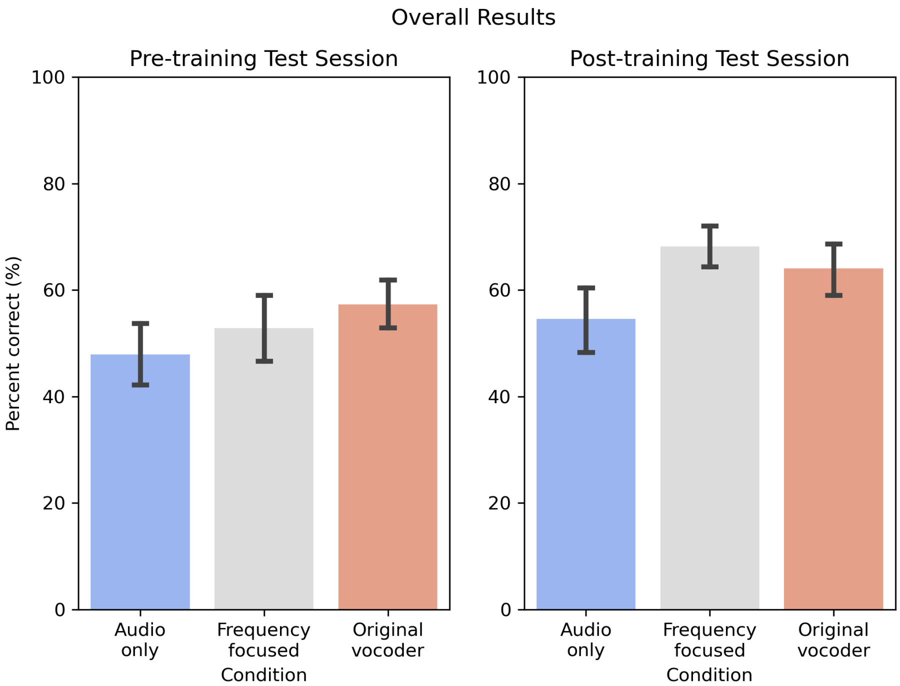

Figure 6.8: Overall speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the distributions.

Group-level speech-in-noise performance for the conditions (audio only, frequency-focused vocoder, and original vocoder) and session type (pre- and post-training) are displayed in Figure 6.8.

A 2-way repeated measures ANOVA was performed with factors "Condition" (Audio only, Frequency focused and Original vocoder) and "Session" (before and after training).

No significant effect was found for either condition ($F(1,5) = 2.56, p = 0.15, \eta_p^2 = 0.34$) or session ($F(1,5) = 4.80, p = 0.08, \eta_p^2 = 0.49$) and no interaction was found between the two ($F(1,5) = 3.46, p = 0.07, \eta_p^2 = 0.41$). For the audio only condition, the average performance was 47.87% before training (between 26.67% and 68.10%, SE=6.14) and 54.54% after training (between 38.10% and 78.92%, SE=6.56). For the frequency vocoded haptics condition, average performance was 52.83% before training (between 24.76% and 67.14%, SE=6.35) and 68.20% after training (between 53.81% and 80.95%, SE=4.30). For the original vocoder haptics condition, average performance was 57.32% before training (between 34.99% and 64.76%, SE=4.67) and 64.04% after training (between 52.50% and 84.76%, SE=5.24).

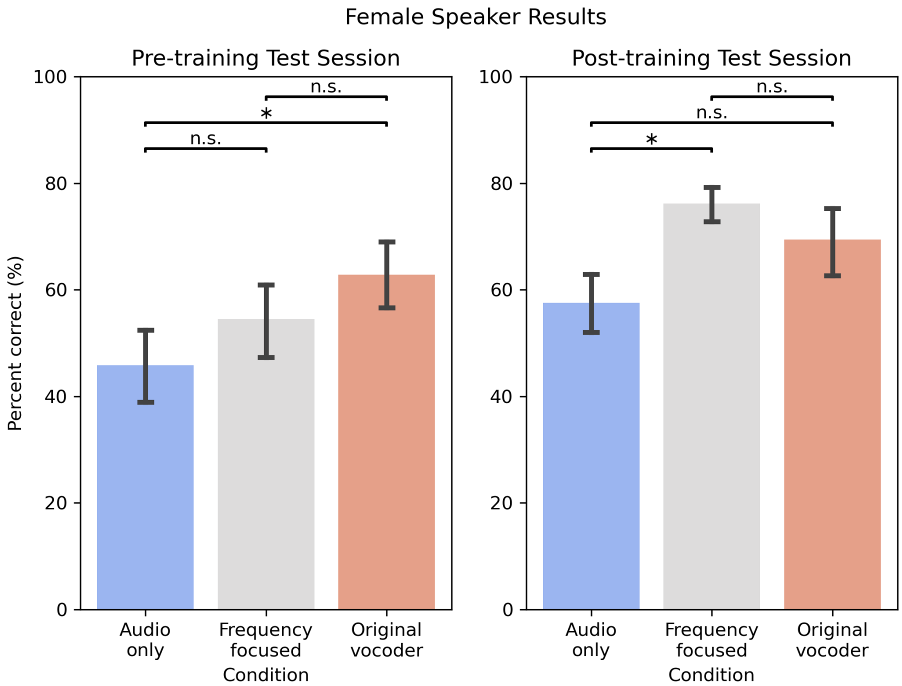

Figure 6.9: Female speaker stimulus only speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the mean.

Two planned post-hoc 2-way repeated ANOVAs were conducted to explore the effect of Condition and Session for speaker genders separately. After correction for multiple comparisons, a significant effect of condition type was found for the female speaker ($F(1,5) = 8.40, p = 0.03, \eta_p^2 = 0.63$). Results are illustrated in Figure 6.9. For the audio only condition, the average performance was 45.82% before training (between 22.22% and 72.69%, SE=7.48) and 57.54% after training (between 41.11% and 80.29%, SE=6.08). For the frequency vocoded haptics condition, average performance was 54.52% before training (between 30.00% and 79.25%, SE=8.02) and 76.16% after training (between 62.65% and 91.11%, SE=3.73). For the original vocoder haptics condition, average performance was 62.82% before training (between 41.65% and 82.22%, SE=6.79) and 69.38% after training (between 42.22% and 86.67%, SE=6.39).

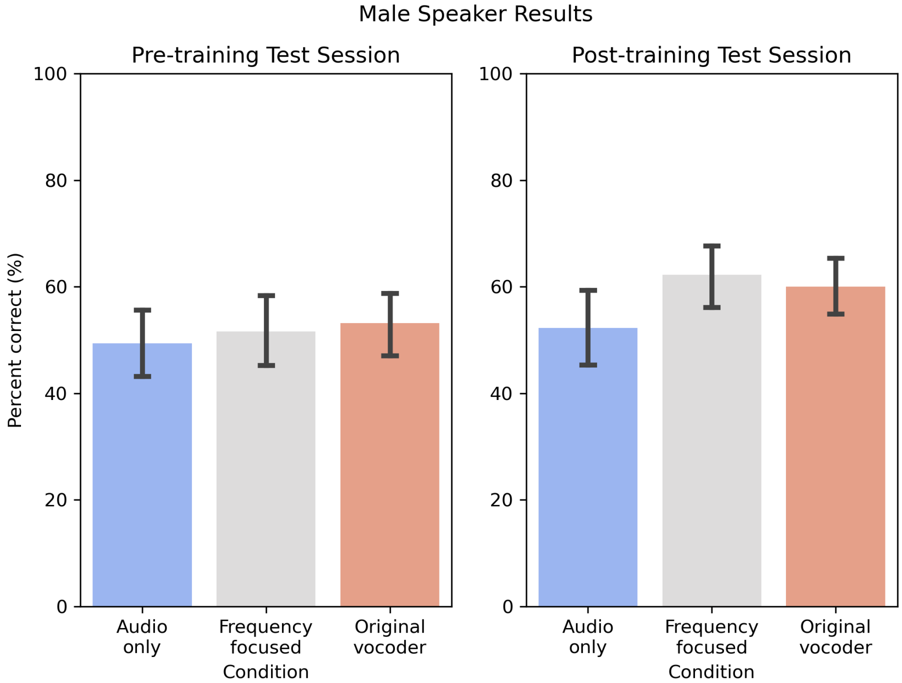

Figure 6.10: Male speaker stimulus only speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the mean.

Post-hoc t-tests were then performed for this data. Before training, the original vocoder performed significantly better than the audio alone condition ($T = -3.75, p = 0.03, d = 5.00$). No significant effect was found between the audio only and frequency focused condition ($T = -1.0, p = 0.36, d = 5.00$), or the frequency focused and original vocoder condition ($T = -1.89, p = 0.24, d = 5.00$). After training the frequency-focused vocoder performed significantly better than the audio only condition ($T = -4.03, p = 0.03, d = 5.00$). No significant difference was found between the audio only and original vocoder condition ($T = -1.6, p = 0.34, d = 5.00$), or the frequency focused and original vocoder condition ($T = 1.1, p = 0.34, d = 5.00$). No effect was found for session type ($F(1,5) = 2.6, p = 0.17, \eta_p^2 = 1.0$) or interaction ($F(1,5) = 1.3, p = 0.31, \eta_p^2 = 0.8$). No further post-hoc tests carried out for these factors. Results for the male speaker data are presented in Figure 6.10

For the male speaker data no effect was found for session type ($F(1,5) = 4.9, p = 0.08, \eta_p^2 = 1.0$), condition type ($F(1,5) = 0.53, p = 0.57, \eta_p^2 = 0.81$) or interaction between the two ($F(1,5) = 1.9, p = 0.21, \eta_p^2 = 0.92$). Again, no further post-hoc tests were performed for these factors.

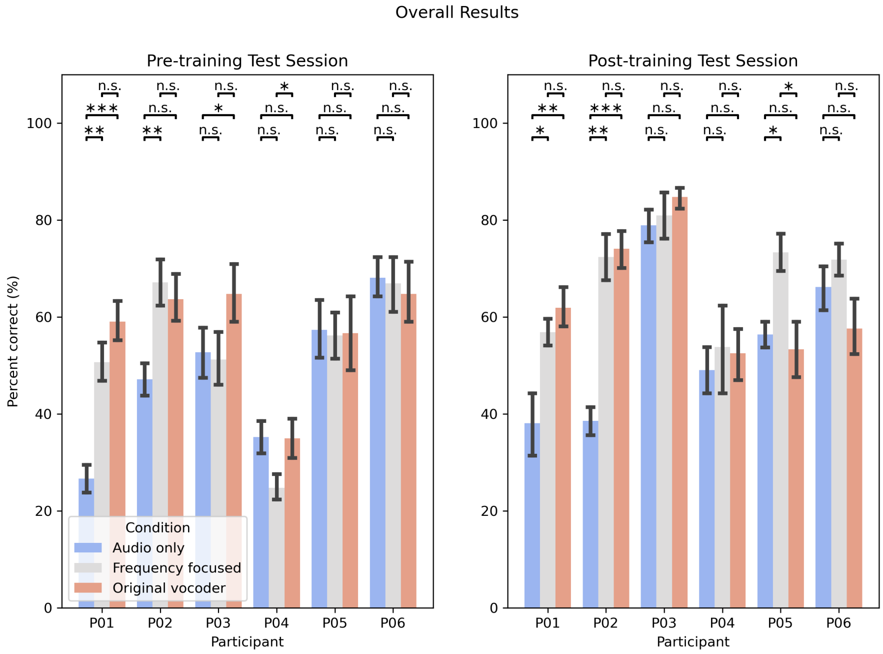

Figure 6.11: Overall individual speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the mean.

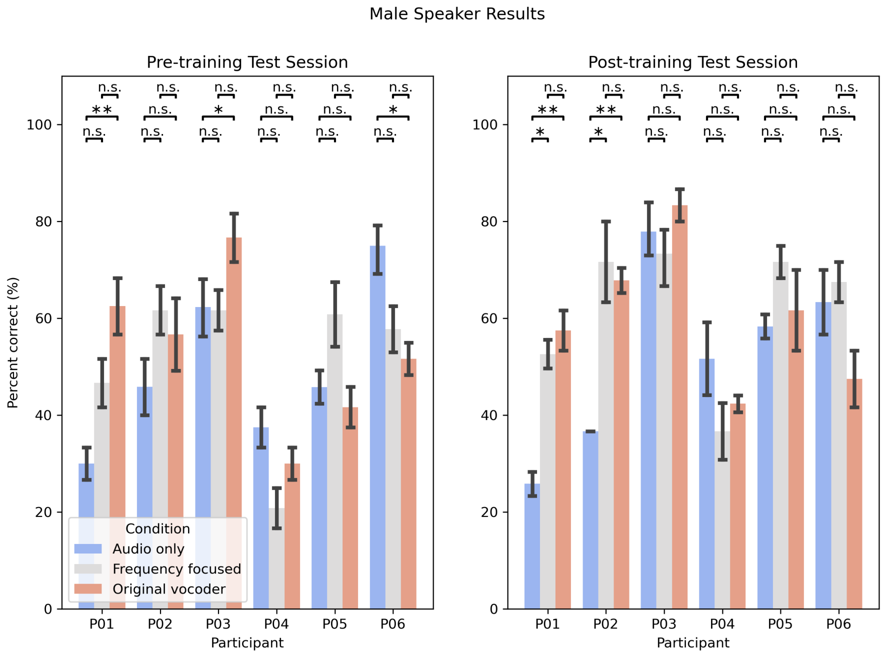

Figure 6.12: Male speaker stimulus only individual speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the distributions.

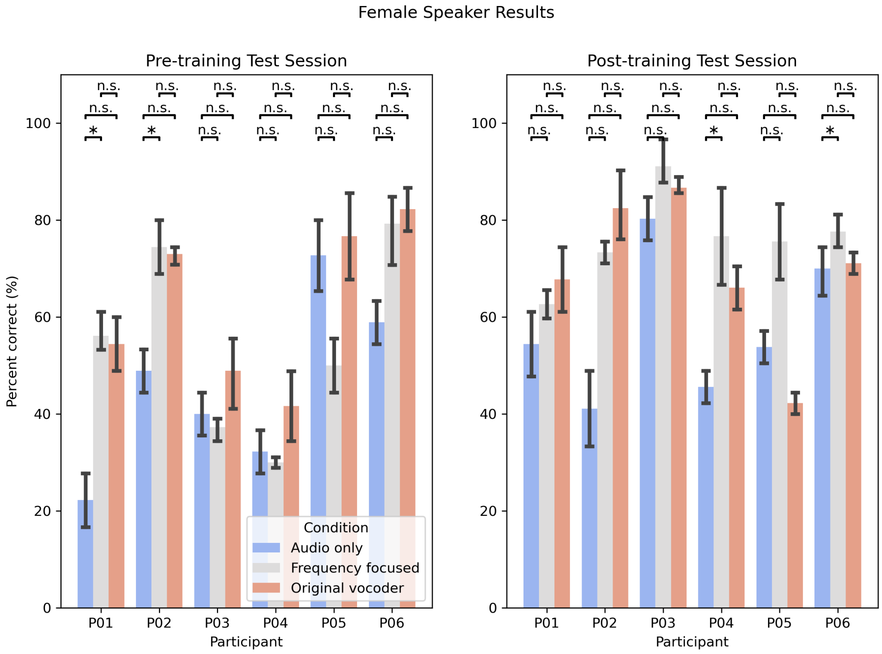

Figure 6.13: Female speaker stimulus only individual speech-in-noise performance for the audio only, frequency-focused vocoder and original vocoder conditions. Bars represent the mean percent-correct across all participants. Error bars show the standard error of the distributions.

Further unplanned *t*-tests were performed to explore the relationship between each condition for each participant. No correction for multiple comparisons was applied to these exploratory tests. Significant differences are illustrated per participant in Figures 6.11, 6.12 and 6.13

Finally, unplanned linear regression analysis was applied to explore the relationship between training session number and participant performance, for each condition. These analyses were first performed using all training and test data. Results are presented for group-level analysis in Figure 6.14. Further analysis of individual participant performance is then illustrated in Figure 6.15. This analysis was also performed using only male or female training/test data, to explore the effect of speaker gender on performance. These analyses are illustrated at the group-level in Figures 6.16 and 6.18, and for individual participants in Figures 6.17 and 6.19.

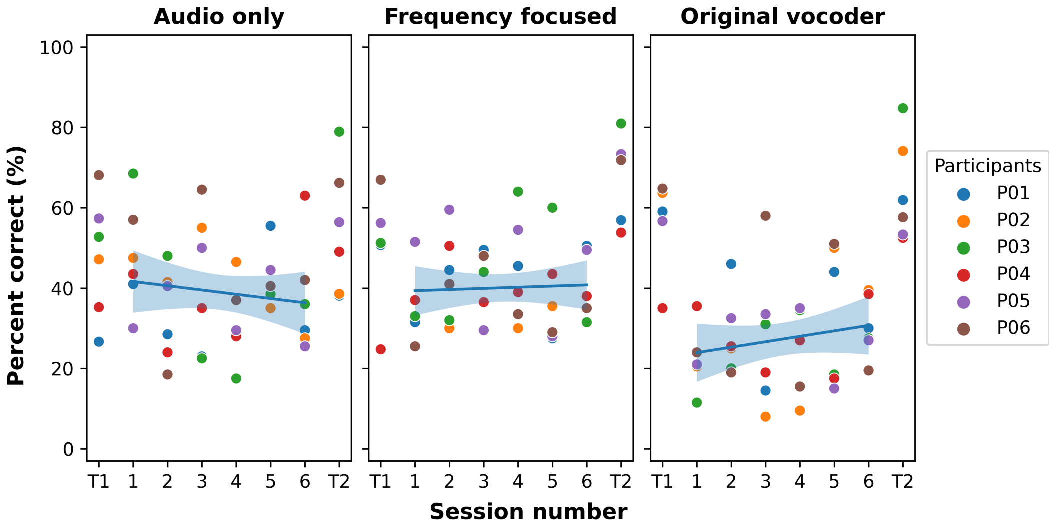

Figure 6.14: Group-level change in performance across training sessions for each condition, using both male and female lists. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Condition	R^2	$F(1,34)$	p	Model (β_0, β_1)
Audio only	0.02	0.69	0.41	42.694, −1.060
Frequency focused	0.00	0.08	0.78	39.061, +0.288
Original vocoder	0.04	1.29	0.26	22.578, +1.355

Table 6.2: Regression statistics for Figure 6.14

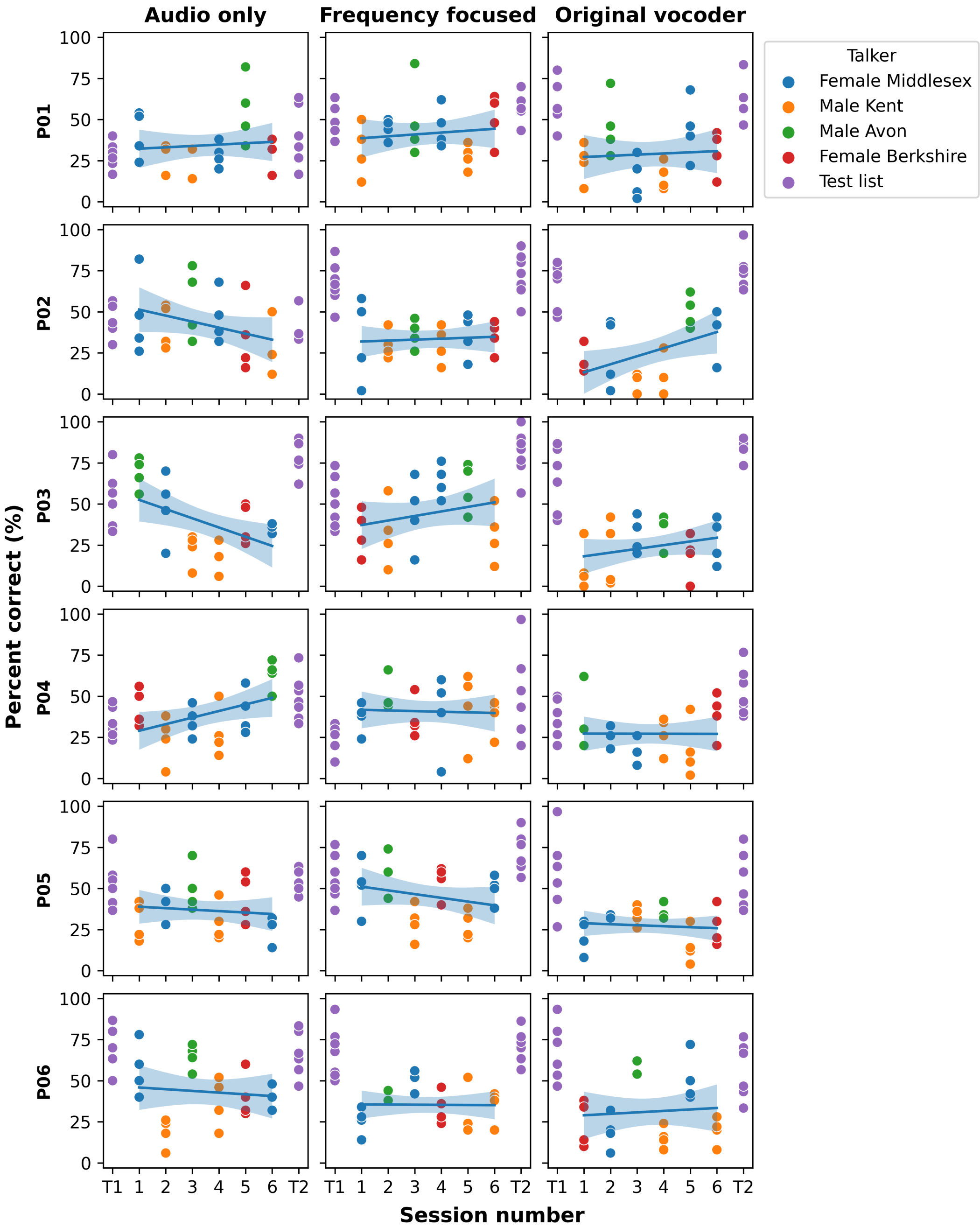

Figure 6.15: Individual participant change in performance across training sessions for each condition, using both male and female lists. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Participant	Condition	R^2	$F(1,22)$	p	Model (β_0, β_1)
P01	Audio only	0.01	0.19	0.67	31.433, +0.829
	Frequency focused	0.02	0.35	0.56	37.500, +1.143
	Original vocoder	0.00	0.10	0.75	26.500, +0.714
P02	Audio only	0.11	2.65	0.12	54.967, −3.657
	Frequency focused	0.01	0.13	0.72	31.333, +0.571
	Original vocoder	0.19	5.23	0.03	8.267, +4.900
P03	Audio only	0.23	6.69	0.02	58.100, −5.600
	Frequency focused	0.06	1.33	0.26	34.433, +2.757
	Original vocoder	0.07	1.66	0.21	15.933, +2.257
P04	Audio only	0.17	4.46	0.05	25.000, +4.000
	Frequency focused	0.00	0.04	0.84	42.100, −0.386
	Original vocoder	0.00	0.00	0.99	27.267, −0.029
P05	Audio only	0.01	0.27	0.61	39.767, −0.886
	Frequency focused	0.06	1.43	0.24	53.367, −2.271
	Original vocoder	0.01	0.22	0.64	29.433, −0.600
P06	Audio only	0.01	0.22	0.65	46.900, −1.043
	Frequency focused	0.00	0.00	0.95	35.633, −0.086
	Original vocoder	0.01	0.14	0.71	28.067, +0.886

Table 6.3: Regression statistics for Figure 6.15

Models reported in the form: $y_{\text{Performance}} = \beta_0 + \beta_1 x_{\text{Session number}}$.

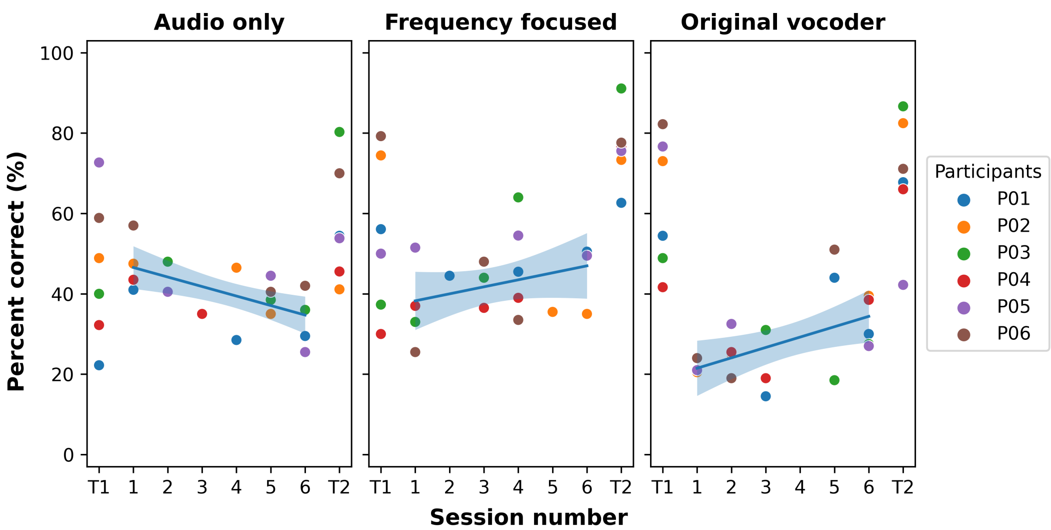

Figure 6.16: Group-level change in performance across training sessions for each condition, using female lists only. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Condition	R^2	$F(1,16)$	p	Model (β_0, β_1)
Audio only	0.36	9.03	0.01	$48.914, -2.367$
Frequency focused	0.11	1.92	0.19	$36.511, +1.742$
Original vocoder	0.27	5.93	0.03	$18.902, +2.581$

Table 6.4: Regression statistics for Figure 6.16

Models reported in the form: $y_{\text{Performance}} = \beta_0 + \beta_1 x_{\text{Session number}}$.

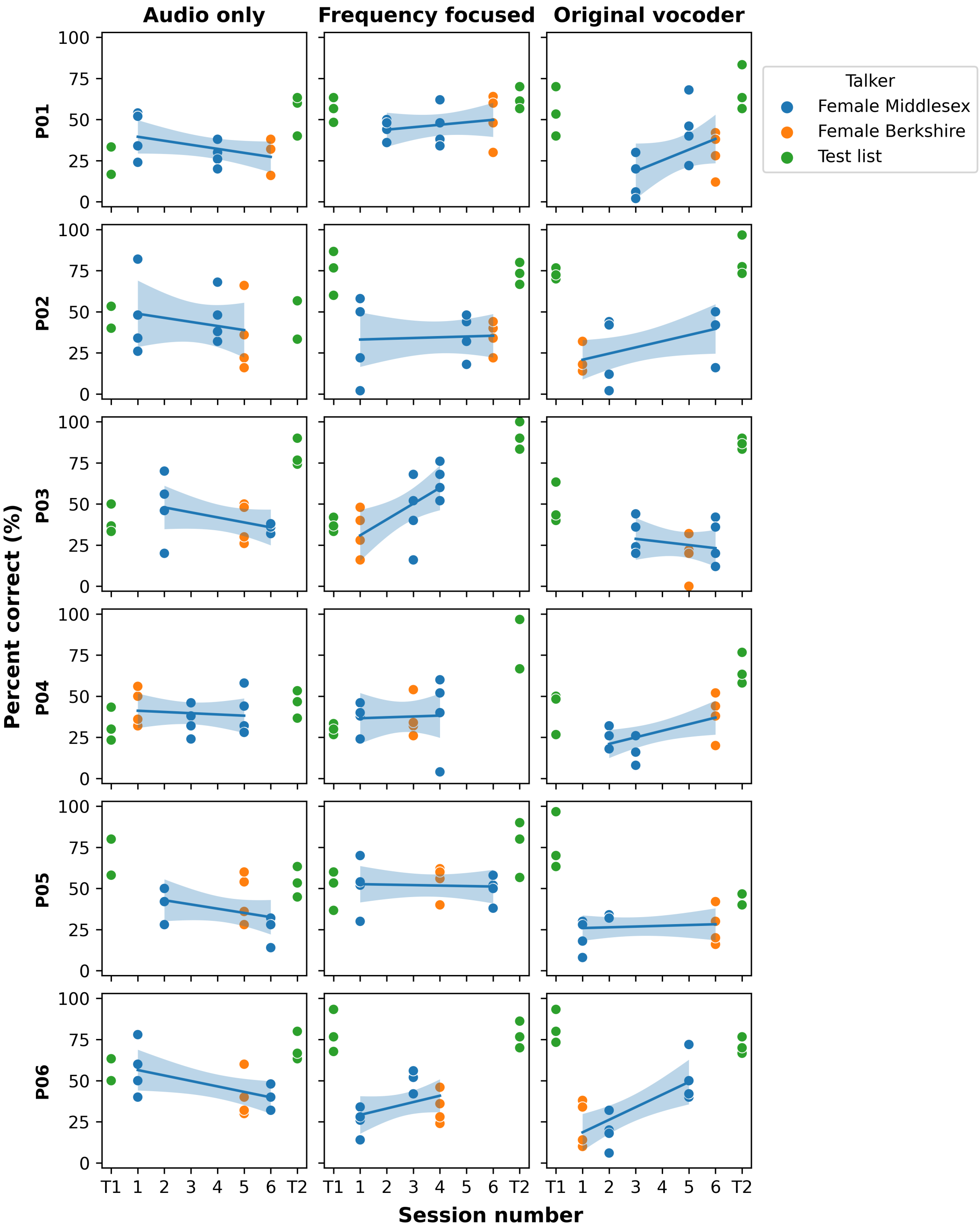

Figure 6.17: Individual participant change in performance across training sessions for each condition, using both female lists only. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Participant	Condition	R^2	$F(1,10)$	p	Model (β_0, β_1)
P01	Audio only	0.21	2.64	0.14	$41.974 - 2.447$
	Frequency focused	0.05	0.56	0.47	$40.833 + 1.500$
	Original vocoder	0.21	2.61	0.14	$-1.000 + 6.536$
P02	Audio only	0.05	0.50	0.49	$51.269 - 2.481$
	Frequency focused	0.00	0.04	0.84	$32.643 + 0.464$
	Original vocoder	0.25	3.40	0.09	$17.083 + 3.750$
P03	Audio only	0.15	1.77	0.21	$54.000 - 3.038$
	Frequency focused	0.40	6.71	0.03	$21.286 + 9.643$
	Original vocoder	0.04	0.38	0.55	$34.500 - 1.893$
P04	Audio only	0.01	0.14	0.72	$41.917 - 0.750$
	Frequency focused	0.00	0.02	0.89	$36.071 + 0.536$
	Original vocoder	0.33	4.97	0.05	$13.000 + 4.000$
P05	Audio only	0.12	1.37	0.27	$48.000 - 2.577$
	Frequency focused	0.00	0.03	0.86	$52.895 - 0.289$
	Original vocoder	0.01	0.12	0.73	$25.440 + 0.464$
P06	Audio only	0.28	3.92	0.08	$59.786 - 3.321$
	Frequency focused	0.17	2.02	0.19	$25.286 + 3.893$
	Original vocoder	0.51	10.60	0.01	$10.923 + 7.654$

Table 6.5: Regression statistics for Figure 6.15

Models reported in the form: $y_{\text{Performance}} = \beta_0 + \beta_1 x_{\text{Session number}}$.

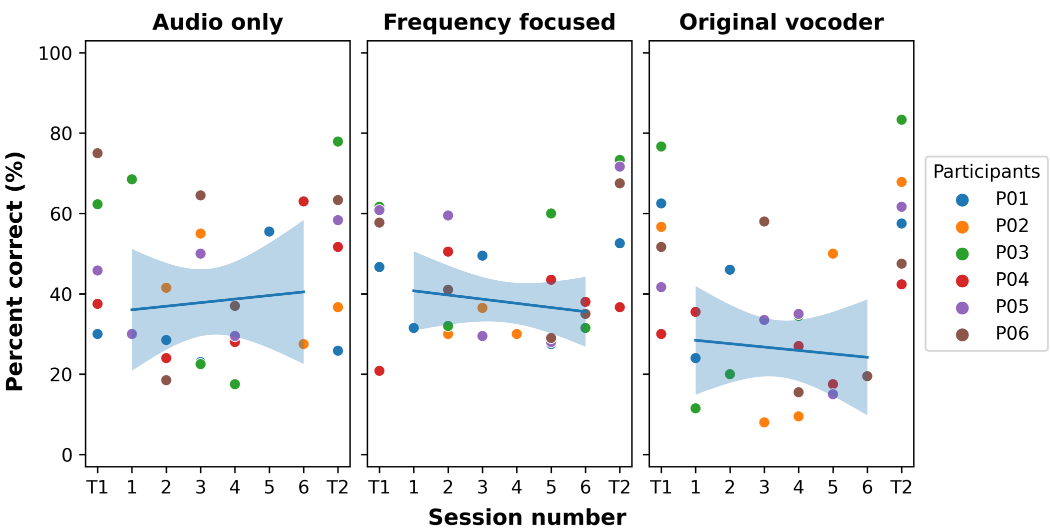

Figure 6.18: Group-level change in performance across training sessions for each condition, using male lists only. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Condition	R^2	$F(1,16)$	p	Model (β_0, β_1)
Audio only	0.01	0.10	0.76	$35.135, +0.889$
Frequency focused	0.03	0.44	0.51	$41.780, -1.038$
Original vocoder	0.01	0.12	0.73	$29.287, -0.847$

Table 6.6: Regression statistics for Figure 6.16

Models reported in the form: $y_{\text{Performance}} = \beta_0 + \beta_1 x_{\text{Session number}}$.

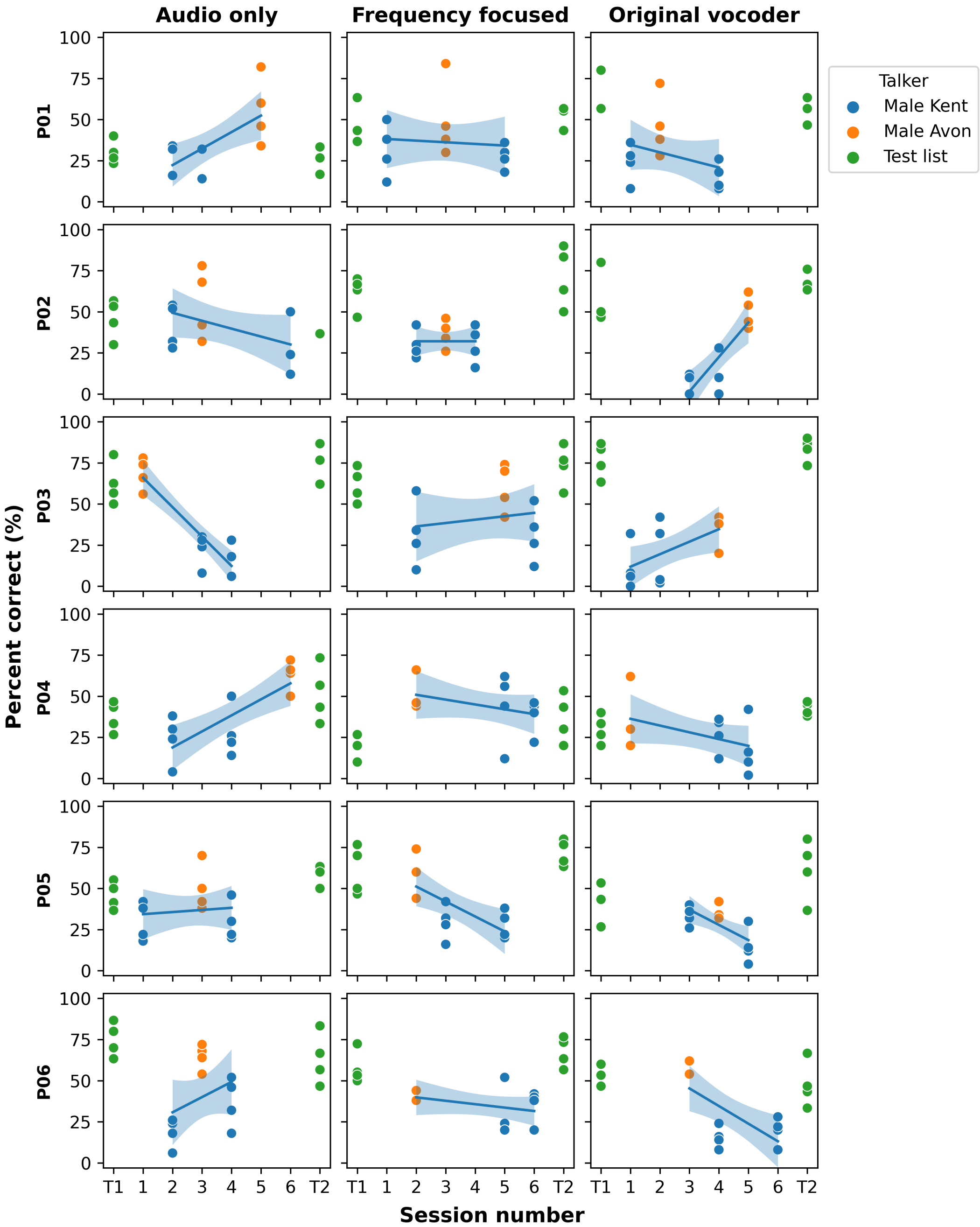

Figure 6.19: Individual participant change in performance across training sessions for each condition, using both male lists only. Points represent all training/test lists completed for all participants. Session T1 represents the initial test session and T2 represents the final test session. All other sessions represent training sessions. The solid line represents an ordinary least-squares regression fitted to the training data. The error bars of this line represent the standard error of the regression line.

Participant	Condition	R^2	$F(1,10)$	p	Model (β_0, β_1)
P01	Audio only	0.44	8.00	0.02	$2.214 + 10.036$
	Frequency focused	0.01	0.09	0.78	$39.167 - 1.000$
	Original vocoder	0.11	1.21	0.30	$39.250 - 4.607$
P02	Audio only	0.19	2.34	0.16	$58.962 - 4.808$
	Frequency focused	-0.00	-0.00	1.00	$32.167 - 0.000$
	Original vocoder	0.65	18.37	0.00	$-61.500 + 21.000$
P03	Audio only	0.83	49.80	0.00	$83.786 - 17.857$
	Frequency focused	0.03	0.31	0.59	$32.250 + 2.058$
	Original vocoder	0.34	5.15	0.05	$4.250 + 7.607$
P04	Audio only	0.57	13.50	0.00	$-0.667 + 9.750$
	Frequency focused	0.12	1.37	0.27	$56.750 - 2.942$
	Original vocoder	0.20	2.53	0.14	$40.385 - 4.115$
P05	Audio only	0.01	0.12	0.73	$33.071 + 1.286$
	Frequency focused	0.44	7.77	0.02	$69.357 - 9.107$
	Original vocoder	0.45	8.16	0.02	$64.833 - 9.250$
P06	Audio only	0.13	1.45	0.26	$12.250 + 9.250$
	Frequency focused	0.11	1.25	0.29	$44.000 - 2.077$
	Original vocoder	0.45	8.03	0.02	$77.429 - 10.714$

Table 6.7: Regression statistics for Figure 6.15

Models reported in the form: $y_{\text{Performance}} = \beta_0 + \beta_1 x_{\text{Session number}}$.

6.4 Discussion

This study found that overall there was no statistically significant effect of either tactile stimulation strategy, but that for specific participants there may be considerable benefits of tactile stimulation. A small but statistically significant effect of frequency-focused vocoder was observed for the female speaker only stimulus after training, compared to audio only condition. This suggests that with training the frequency focused algorithm may have the potential to improve speech-in-noise performance for cochlear implant users, but that this benefit may be dependent on the acoustic properties of the attended speaker. A small but statistically significant effect was also found for the original vocoder in comparison to the audio only condition. This was only found before training on the female only speaker stimulus. This may suggest that the training regime used for this experiment was not suitable for the original vocoder strategy. No statistically significant difference was found between the signal-processing strategies, which suggests that there was no demonstrable group-level advantage of the new frequency-vocoder approach over the original vocoder approach.

On an individual level, there were considerable improvements for specific users (although the lack of correction for multiple comparisons should be considered with these exploratory analyses). Improvements of up to 32.38% better than audio only were found for the original vocoder and up to 24.04% for the Frequency-focused vocoder. These results suggest that although the proposed haptic stimulation method may not provide the expected performance for all users, certain user may see considerable performance benefits from this technique.

For the test data, the lack of a group-level effect in this work contrast previous findings, where statistically significant effects have been found for similar signal processing strategies to the original vocoder condition. For example, Fletcher et al. (2018) observes an improvement of 10.8% words correct scores on average (with a standard error of ±2.2%) for a similar cohort of users. In fact, an almost identical average performance benefit was found for the original vocoder condition (9.5%) but with a higher standard error of ±6.5%. This high-variance in results suggests that the lack of participants (as a direct result of the COVID–19 pandemic, see the Statement on the impact of COVID-19) may have precluded the observation of a robust effect in this study, as was observed for the method presented by Fletcher et al. (2018). Similar observations may also be made when comparing results to those of a study in real cochlear implant users by Fletcher et al. (2019), which found improvements of around 8.3% ($SE = 2.5\%$). On an individual level the extrema for the current study also suggested promise for the proposed approach. For the original vocoder, the best participant scored 32.38% better than the audio only condition (highest scores for the previous studies were 17.8% and 21.8% respectively), with no score significantly worse than the audio alone condition. The Frequency focused strategy also performed better than the audio alone condition, scoring 24.04% for the best user. However, the worst user scored 10.48% lower than the audio only condition suggesting that this vibration may have distracted participants from the task.

In addition to the test data, analysis was performed for the training session data. Analysis of performance in the training sessions suggested that participants did not improve over time with the proposed training regime. For each of the conditions, no statistically significant positive trends in performance were observed across sessions. However, when analysing the female trials alone, a significant positive increase of 2.58% in performance per-session was observed for the original vocoder condition. This suggests that although overall the training sessions were largely ineffective, some users may have benefited from training on female talkers, when using the original vocoder. Again, a lack of correction for multiple comparisons should be considered for these exploratory analyses.

For the analysis of individual training performance, similar results were observed to group-level analysis. A significant improvement of 4.9% per-session was found for participant 2 in the original vocoder condition. A negative effect of -5.6% was also found for participant 3. However, for the majority of participants no significant trend was observed, and the significant results would likely not survive correction for multiple comparisons. This suggested that, overall, the majority of participants did not benefit from the training regime.

In the female- and male-only analysis of individual the effects of training was mixed. Some participants showed improvements in performance for certain conditions, but benefits were not consistent across either participants or conditions. In certain cases, negative effects were also observed, suggesting that participants actually got worse over time. For the female-only analysis, participant 5 showed significant improvements of 7.6% per-session for the original vocoder, and participant 3 showed improvements of 9.6% improvement for the frequency focused condition. Again, the majority of participants did not show significant trends across sessions. However, for the Male only analysis both significant positive and negative effects were observed (without correction for multiple comparisons). Whilst significant benefits of up to 21% per-session were shown for 3 participants in at least one condition, the other 3 participants showed a significant negative effect of training up to 17.9% across sessions. It should be noted that for the male and female only analysis, the lack of training data for each gender on every session may alter the per-session change in performance values. Therefore, the length of time between sessions should also be considered when interpreting these values (illustrated in Figures 6.17 and 6.19).

There are a number of limitations in this study that should be considered. As mentioned, one of the most detrimental limitations of the study was the lack of power in statistical analysis due to the low number of participants. This limited the applicability of any findings to the development of a haptics intervention — Further assessment of these techniques would be needed to more conclusively understand the effects of a spatial mapping of audio to tactile stimulation for speech-in-noise performance improvement. It should also be noted that the frequency focused vocoder's adaptive filtering does not provide the same broader spectral features that the original vocoder provides. This design aimed to provide the resolution necessary to portray the sub-semitone pitch changes that encode segmental (Oxenham, 2008; David et al., 2017) and suprasegmental (Banse and Scherer, 1996; Murray and Arnott, 1993; Most and Peled, 2007; Peng et al., 2008; Meister et al., 2009; Xin Luo et al., 2007) pitch cues that are lacking in the original vocoder approach. As the broader speech envelope is thought to provide key cues for speech intelligibility (Shannon et al., 1995; Drullman et al., 1994) this may be a reason for the lack of an increased effect over the original vocoder strategy. Future work should consider the balance between the representation of fine-grain pitch cues and broader spectral envelope cues in haptic signal processing strategies. In addition only simulated cochlear implant users were tested, with an average age considerably younger than that of an average cochlear implant user. Previous research in this area has shown comparable results between cochlear implant users and equivalent normal-hearing listeners using a cochlear implant simulator (Fletcher et al., 2018; Fletcher et al., 2019) but further investigation is needed to confirm this.

There were also significant limitations of the training regime. For example, the use of both male and female materials with multiple accents in both training and testing may have had detrimental effects on participant performance. Analysis of the training data suggested little benefit of this regime and informal reports from some participants suggested that they struggled when switching between listeners, as this required them to adjust to the change in haptic stimulation and in the cochlear implant simulator (possibly as a result of changes to the acoustic and linguistic properties of the speaker). More focused training on individual speakers and a reduction in the number of speaker types may benefit future studies, given the limited amount of time available for training participants. Future studies may look to address this through development of real-world usable devices, allowing users to train more intensively with the device, in more ecologically valid scenarios and for longer periods of time. Finally, the effect of training performance on testing performance was not evaluated in this study. Future work should look to quantify the relationship between the training regime and the outcome performance. This would provide further insight into the generalisation of the training material to unseen speakers and noisy environments.

There are also a number of areas that future work could explore to maximise the benefits of haptic stimulation for the improvement of speech-in-noise performance. This study has suggested that results may vary considerably between users, with some seeing greater benefits than other. Further work is needed to determine the users that will see the greatest benefits from haptic stimulation, as well as the issues that prevent other participants from seeing significant benefits. This study also aimed to present a device that could be readily translatable to a real world intervention. However, further development is needed to achieve this - For example, the use of ERM motors may not have optimally presented the amplitude envelopes to the participant, given the co-modulation of frequency and amplitude. Whilst these motors have been shown to be effective for presenting simplistic pitch cues (Fletcher et al., 2020c), this complex interaction of both amplitude and frequency may not provide the most intuitive spatial mapping of amplitude envelopes across the skin. Other motor types such as LRAs or Piezo actuators should be evaluated to determine the optimal presentation of the speech cues. Other areas that should be explored in greater detail include the power consumption of motors, use of wireless technologies to reliably connect the device to cochlear implant with minimum latency, and the miniaturisation of electronics and signal-processing to fit on a wrist worn device. Advances in these areas are critical to moving the technology towards a more feasible real-world intervention for cochlear implant users. This study also adapted the frequency-based mapping of previous vocoder-based studies to a spatial mapping across the wrist. This type of mapping may also show further benefit for similar areas such as sound localisation (Fletcher et al., 2020a; Fletcher et al., 2021a; Fletcher et al., 2021b) and speech in spatially separated noise (Fletcher et al., 2020b).

This study has presented a new method for improving speech-in-noise performance for cochlear implant users. Results showed minor benefits for select stimulus, but not significant overall effect. However, for individual participants considerable performance improvements were observed, suggesting that this method may be viable for providing considerable benefit to speech-in-noise performance for some cochlear implant users. Suggestions have been made to address the issues that may potentially have limited the performance increases observed in this study.

6.5 Contributions

All sections of this chapter were designed and implemented by the author Samuel Perry and supervisor Mark Fletcher.

Chapter 7

Discussion

The research presented aimed to address issues with speech-in-noise performance for the hearing impaired by focusing on two challenges. The first, is how to effectively assess and diagnose a speech-in-noise performance issue. The second, is how to best restore performance, typically achieved by fitting an intervention.

Chapters 3 and 4 focused on assessment of speech-in-noise performance, investigating the optimisation of the TRF that may be used for both clinical assessment and in-lab analysis of speech-in-noise performance. Previous literature has suggested that the proposed TRF method may be adapted to form an effective technique for assessment of speech-in-noise performance (Vanthornhout et al., 2018; Decruy et al., 2018; Lesenfants, 2019). However, these studies have only demonstrated limited predictive performance and required large datasets of evoked responses. Chapters 3 and 4 investigated the benefits of modifications to both the TRF regression model type, and to the selection of input audio features to be reconstructed by the model. Any benefits of these optimisations could then be directly applied to the methods for analysing speech-in-noise performance described in previous literature.

In Chapter 3, a minor benefit of using an ElasticNet model was found over current models for reconstructing TRFs. This suggested that additional model regularisation could provide modest performance improvements, and that overfitting was not a primary issue in model training. However, the ElasticNet model offered additional benefits, such as implicit feature selection. This may be of benefit for understanding the contributions of electrodes/features to the trained model and for selection of the optimal subset of electrodes. These enhancements may be used in the translation of the TRF to a clinical method. Chapter 4 presented optimisations of the input features, reconstructed using the elastic-net TRF model from Chapter 3. In this chapter the loudness model was reconstructed with significantly higher accuracy than the traditional amplitude envelope model used in previous literature. This suggested that a loudness model based feature was better represented in cortical evoked potentials. These results highlighted this type of perceptually motivated feature as a suitable alternative for use in evaluation methods such as those presented by Vanthornhout et al. (2018). Combined, Chapters 3 and 4 suggest that modifications to the TRF method can improve its stimulus reconstruction performance. These modifications are readily adaptable to proposed methods for assessing speech-in-noise performance in the lab. Additional work is needed to provide a suitable means for assessing speech-in-noise performance in clinic. This is discussed further in Sections 7.1 and 7.3.

Chapters 3 and 4 focused on development of a new multi-sensory intervention to improve speech-in-noise performance for cochlear implant users. Through use of a new haptic wearable device, Chapter 5 demonstrated considerable benefit to pitch discrimination for normal-hearing participants using a cochlear implant simulator. This finding suggests that pitch cues (which are important features of speech recognition in noise) can be transmitted effectively with an appropriate audio-to-haptic signal processing strategy and spatial array of haptic motors. Additionally, the robustness of the method to noise suggested that the approach could be highly effective in situations which cochlear implant users currently have great difficulty.

Chapter 6 modified the approach presented in Chapter 5 to assess the improvements to speech-in-noise performance in normal-hearing participants using a cochlear implant simulator. Results showed a small benefit of the proposed frequency-focused vocoder tactile stimulation method for female speakers after training, but showed highly significant benefits for some individuals. However, despite the magnitude of differences between conditions following that of previous research (Fletcher et al., 2018; Fletcher et al., 2019), a lack of statistical power may have resulted in no overall effect of the haptic stimulation. In addition, results suggested that this method may be particularly effective for certain users, and that further optimisations to both the signal-processing strategy and training regime may be factors in the lack of a more robust effect.

These studies contribute to a growing literature of evidence, suggesting that haptics may provide considerable benefit to speech-in-noise performance for cochlear implant users. The significant improvements in pitch discrimination demonstrated in Chapter 5 provide the groundwork for providing fundamental acoustic cues required for understanding speech in noise.

Overall, the results of this thesis have contributed to the development of two methods that have application in improving speech-in-noise performance for the hearing impaired. These show promise both as independent methods, and also in combination, as is discussed in Section 7.3.

7.1 Future development of the TRF method

The presented results demonstrate that features thought to be important for speech-in-noise performance, such as perceptual correlates to the amplitude envelope, can be reliably reconstructed using the TRF method. However, these results do not directly assess the sensitivity of the measure to speech recognition in background noise. As a result, the presented TRF analysis method can be thought of as between levels 2 ("Technology concept and/or application formulated") and 3 ("Analytical and experimental critical function and/or characteristic proof-of-concept") of the Technology Readiness Level (TRL) standard, as defined in ISO 16290:2013 (Space Systems. Definition of the Technology Readiness Levels (TRLs) and Their Criteria of Assessment 2013). The work in this thesis has supported the formulation of an application for the TRF method (TRL 2 — i.e. as an objective measure of speech-in-noise performance, and analysis method for assessing multi-sensory integration) and has provided lab-based experimental results that contribute to a growing literature, demonstrating the limits of the TRF method (TRL 3). Whilst the performance requirements are broadly defined for the technology (as outlined in Chapter 2). In order to develop a full proof-of-concept in-lab and clinical speech-in-noise measure (TRL 3), further work is needed to assess the correlation of this method with behavioural speech-in-noise measures. Future work should compare the predictive performance of the proposed method to those presented by Vanthornhout et al. (2018) and Lesenfants (2019). This would provide insight on the degradation of reconstruction performance with increasing noise levels, a crucial step in the development of an objective measure of performance.

For further development of the method, particularly as a clinical measure, factors such as the effect of attention, hearing impairment and age on the method's performance should also be assessed. The effect of attention has been investigated by Vanthornhout et al. (2019). Results suggest that attention does affect the TRF for amplitude envelope reconstruction but that without attention, reconstruction performance still correlates well with the SNR of the stimulus. This suggests that the method may be implementable for cohorts such as infants, who may not actively attend the stimulus.

The effect of age and hearing impairment on TRF amplitude envelope reconstruction are measured by Decruy et al. (2018) and Decruy et al. (2020). An increase in reconstruction performance is observed with increasing age and with increasing severity of hearing loss, despite a decrease in intelligibility scores. This should be considered in particular for methods such as those presented by Vanthornhout et al. (2018), where reconstruction performance is expected to degrade with reduced speech-in-noise performance. Further work is also needed to understand the underlying mechanisms that drive these increases in reconstruction performance. The areas outlined would contribute to the proposed method reaching TRL 3.

For a more practical implementation of the measure (as required for TRLs 4–5 — "Component and/or breadboard functional verification in laboratory environment", "Component and/or breadboard critical function verification in a relevant environment"), further work is needed to reduce the number of electrodes used in testing (Montoya-Martínez et al., 2021), and to determine the optimum stimulus type Verschueren et al. (2020) and analysis frequency bands (Zuk et al., 2021; Synigal et al., 2020). This would be needed to reduce test time in clinic and optimise measure's sensitivity to speech-in-noise performance. For reduction of test time in particular, model training on smaller datasets will also be necessary to make the method viable in clinic. There is potential to use techniques such as sequential testing to limit the test time (Chesnaye et al., 2019). When building models with reduced quantities of data, regularisation such as that provided by ElasticNet may offer additional performance improvements that were not shown in this thesis.

Work will also be necessary when considering this method for use with cochlear implant users. Predominantly the issue of the stimulus artefact will need to be managed to avoid confounds in reconstruction performance. Artefact reduction methods such as that presented by Somers et al. (2019) have shown promise in reducing these artefacts, which may making this method viable for cochlear implant users. For cochlear implant users there may also be the additional potential of recording EEG data directly from cochlear implant electrodes, eliminating the need for external electrodes (Somers et al., 2021)

TRLs 6–8 ("Model demonstrating the critical functions of the element in a relevant environment", "Model demonstrating the element performance for the operational environment", "Actual system completed and accepted for flight ("flight qualified")"), work will focus on validation of the method, and any accompanying prototype hardware in real-world clinical environments. The proposed implementation would require validation on multiple clinical populations with a variety of hearing losses, ensuring accurate estimations of speech-in-noise across all relevant cohorts. Practical issues, such as hardware and software integration with standard clinical setups will be necessary. For this, an all-in-one device, including electrodes and a synchronised audio and EEG recording module may need to be developed. Design of the device will also be necessary, to provide intuitive feedback to clinicians, given the limited time available in typical clinical environments. An alternative research interface may also be required, to provide researchers with the increased signal processing parameter control necessary for in-lab experiments. Appropriate interpretable measurement values must also be produced by the device, such that they can be easily translated into intervention fitting parameters. Measurement data should be collated, within reasonable limits, to allow for assessment of the method's performance. This will allow for informed development of further method updates and optimisations. Finally, the development and roll-out of the finalised software and hardware package to relevant clinics and labs will signify completion of TRL 8.

TRL 9 (“Actual system “flight proven” through successful mission operations”) will be completed on acceptance of the method into standard clinical and research practices.

In addition to the presented method’s application as a measure of speech-in-noise performance, it may be of use in areas such as development of a closed loop fitting system for acoustic interventions or haptic interventions (discussed further in Section 7.3), where the intervention dynamically optimises the intervention based on neural responses measured on-the-fly (Guger et al., 2021a; Guger et al., 2021b, p.53-64, p.95-104). It may also be used for attentional decoding (Bednar and Lalor, 2020; Teoh and Lalor, 2019; O’Sullivan et al., 2019) to steer gain control and noise reduction algorithms in hearing aids (Aroudi and Doclo, 2020; Das et al., 2018; Geirnaert et al., 2020; Van Eyndhoven et al., 2017). This technology could be further applied to cochlear implants and haptic interventions. Finally, the method may also be adapted to explore the use of alternative features, such as pitch (Teoh et al., 2019) to assess degradation of prosody recognition in noise. This may also be applicable analysing the neural underpinnings of music perception in the hearing impaired (Zuk et al., 2021).

7.2 Future development of haptic interventions

As with previous Section 7.1, future development towards the realisation of a real-world haptic intervention for hearing impaired listeners will be described in the context of the TRL scale. Currently, this technology is between TRLs 3 and 4, as the research presented in Chapters 5 and 6 have begun the process of developing early prototype devices, based on previous lab-based conceptual research by Fletcher et al., 2018 and Huang et al., 2017. In order to translate the results presented in these chapters to provide real world benefit, it will first be necessary to verify the presented results for actual cochlear implant users. Beyond this, focus should be placed on development of a suitable wrist-worn haptics device and adaption of the signal processing strategies to function on its embedded hardware. Progress has been made in this area throughout this thesis. The development of real-time signal processing strategies presented could be adapted to run on low-power embedded hardware, progressing from the offline approaches taken by Fletcher et al. (2018) and Fletcher et al. (2019). The presented haptic devices could also be more readily translated to a real-world intervention than the large shakers used by Fletcher et al. (2018) or finger mounted devices used by Huang et al. (2017). Given the limitations of the devices presented in Chapters 5 and 6, such as the size of the forearm mounted device in Chapter 5 and the lack of clear benefits shown for the device presented in 6, significant modifications to the device design will be necessary. To address the size and power requirements of the ERM motors used in this Thesis, LRA or Piezo actuators may be more appropriate for future designs, as discussed in Section 6.4. The use of ERMs is also limited by the inherently linked frequency and amplitude of the motors (Fletcher and Verschuur, 2021). Use of alternate motors without this limitation may yield better performance outcomes than those reported in Chapter 6. Any modifications to current device designs will also need to be validated in relevant clinical populations.

Validation of the prototype device will mark the transition of the technology to TRL 4. Further development of the device, in TRLs 5–7, will require consideration for practical technical challenges, such as implementation of effective audio streaming from behind the ear devices, more stringent power requirements and limited signal processing power on the device. Many of these issues may be solved with existing hardware (such as Bluetooth LE for audio streaming). For the power requirements a target of at least 14 hours battery life is recommended (Fletcher, 2021a), allowing users to charge their device overnight. Technologies such as lithium-ion batteries and low-power haptic motors such as piezo electric motors or linear resonating actuators may also allow for greater longevity in this area.

As discussed in previous Chapters 5 and 6, adapting signal processing methods already implemented for cochlear implants and hearing aids (such as automatic gain control, mapping strategies and noise reduction algorithms) may provide additional performance benefits without the need for considerable adaption to work with haptics devices. These methods will also already be optimised for small hardware form-factors, requiring less tuning to optimise power requirements. Further work may also explore the possibility of tuning the signal processing strategy based on that of the user's pre-existing hearing assistive device. Providing sound information that is poorly transmitted via this device on a case-by-case basis may offer further benefits to speech in noise performance (see Fletcher and Verschuur (2021) for an in-depth review of this area).

In TRLs 8–9, further practical areas should be considered for maximum uptake of haptic interventions, including device aesthetics, compactness, discreetness, and comfort. As discussed in Section 2.4, the advent of compact and lightweight haptics drivers and motors, as well as innovation in chip design (including large reductions in size and future potential for flexible microprocessors (Biggs et al., 2021)) allow for more attractive modern designs than were possible during the development of the original haptic aids of the 1980s. Future work may include the development of a compact wrist-worn device such as that illustrated in Figure 7.1

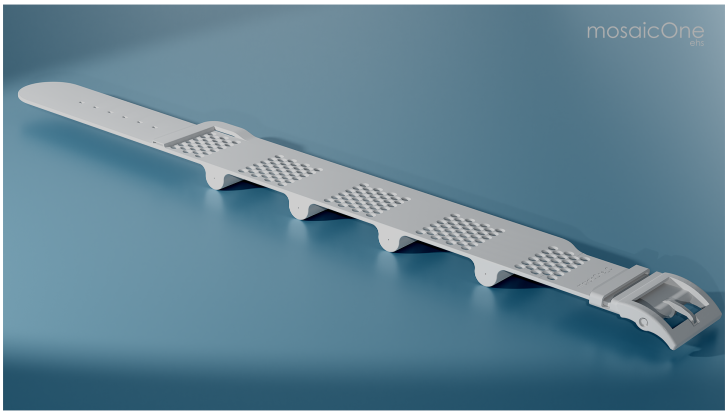

Figure 7.1: Render of a prototype haptic device design. The design includes 4 haptic motors, encased in a flexible, 3D printed wrist strap.

When developing a real-world haptics device, the ease of use for both patient and clinician will be of paramount importance to ensure maximum uptake. This will require device fitting, which will need to be both fast and intuitive for clinicians. Technologies such as automatic motor calibration were used in Chapter 6 to account for physical differences between motors (such as motor pressure and variations/tolerances in motor manufacturing). However, this calibration procedure does not fully account for physiological and perceptual differences between participants and stimulation sites. Chapter 6 also presented a method for tuning a device based on the participant's ability to discriminate between motors. This could also be extended to use additional information, such as absolute thresholds to further maximise the dynamic range, similar to the use of PTA for other hearing interventions. Further investigation is needed to ascertain the benefits using measures such as the participant's tactile detection and discomfort thresholds for this purpose. For this work critical evaluation of the limits of current standards to measures tactile thresholds may also be necessary, as highlighted by Perez et al. (2010).

The presented studies and previous literature for improvement of speech in noise performance have also not yet explored the effects of arm/body movement. Studies have shown that tactile threshold shifts occur as a result of movement (Juravle et al., 2013) and so this may have an impact on the real-world performance, in comparison to the growing literature of primarily in-lab experiments, where stimulation sites are static.

Future experiment protocols should also aim to better represent the challenges faced by cochlear implant users in the real world. Whilst the presented studies used standardised materials to assess speech-in-noise performance, these short, isolated sentences may not represent the challenges (such as additional working memory requirements and cognitive load) faced when attending to longer running speech encountered in everyday life (Larsby and Arlinger, 1994; Hygge et al., 1992). The impact of lip reading will also need to be assessed in combination with haptic stimulation to understand the benefit that haptic stimulation provides when these visual cues are available. Finally, the impact of age on speech-in-noise benefits should be formally assessed, given the changes in haptic sensitivity that occur with increasing age (Lévêque et al., 2000).

The proposed approach may also be adapted to provide haptic benefits for spatial hearing. Cochlear implant users struggle to locate sounds, leading to impaired threat detection and segregation of sound sources (Dorman et al., 2016). Applying the proposed haptic enhancements to both wrists may also aid in this area. The signal processing strategy may offer an enhanced method, providing additional pitch cues to previously proposed methods for improving sound localisation, such as those presented by Fletcher et al. (2020a), Fletcher et al. (2021a), and Fletcher et al. (2020b). This would allow for spatial cues (such as inter aural level differences) to be presented via haptic stimulation in addition to the pitch cues provided in the proposed methods. This may aid in perception of speech features such as prosody, and may be applicable for separating multiple instruments based on both harmonic and panning cues when listening to music (as reviewed in Fletcher (2021b)).

Music perception is a further area that has not yet been fully explored for improving cochlear implant user's listening experience. Evidence suggests that vibration can influence the quality of live concert experiences (Merchel and Altinsoy, 2014) and ability to synchronise dancing to music (Tranchant et al., 2017; Shibasaki et al., 2016). Tactile stimulation devices have also shown improvements in timbre and instrument discrimination (Russo et al., 2012). The presented research, particularly of Chapter 5, may serve as the groundwork for developing a haptics for music. Further challenges include applying effective source separation to process individual instruments, extraction of relevant sound features and intuitive mapping of these to haptic stimulation. An in depth review of this area of haptics research is provided by Fletcher (2021b).

Finally, extra features may be added to a haptic device to aid users with every-day activities. The device could be connected to many other smart devices within the Internet of Things to provide additional functionality, such as alarms, smartphones and televisions. Development of these features would be particularly viable should the developed device be controllable via a smartphone, as any additional processing and communication between smart devices could be outsourced to the smartphone.

7.3 Future work for combining haptics and evoked responses

The research in this thesis has proposed two individual methods (the TRF neuroimaging method and a haptics based intervention) that, with further development, may be used for the assessment and improvement of speech-in-noise performance. As discussed in previous sections, both methods are at similar TRLs, with potential to develop each into clinically viable tools. At their current stages, further benefits may be possible by combining these methods. There are 3 main areas in which these methods could be effectively combined to improve speech-in-noise performance, and more broadly in areas such as sound localisation and music perception. The first is in using the TRF for analysis of the neural mechanisms that underlie the demonstrated haptic benefits. Numerous previous studies have shown extensive connections between somatosensory and auditory neurons at all stages along the auditory pathways (Meredith and Allman, 2015; Basura et al., 2012; Kanold and Young, 2001; Kanold et al., 2011; Shore et al., 2000). However, little is known about at which level haptic integration occurs along the auditory pathways to result in the improved speech-in-noise performances observed. The TRF may serve as a method for analysing this integration, by comparing TRFs trained on haptic stimulus alone, audio-haptic stimulus and audio stimulus alone. With this method, an improvement in feature reconstruction performance for the combined audio-haptic condition (in comparison to the audio only and haptic only conditions) may suggest integration at that level of the auditory pathways. This approach has already been used by Riecke et al. (2019), who found enhanced cortical tracking of the speech envelope when audio and haptic stimulation were provided together. Similar studies, such as that by O'Sullivan et al. (2021) have also already been conducted to explore the integration of audio and visual stimulus of lip reading. Further work may also look at integration at lower levels of the auditory pathway, the effects of training on neural representations and for features such as the loudness model presented in Chapter 4. This could be carried out in the context of additional research at TRL 1–3, which could be used to inform device designs and to optimise signal processing strategies when developing a haptics device (presently between TRLs 3–4, as discussed in Section 7.2).

As mentioned in Section 7.1, the TRF method has the potential to be used for automatic steering of intervention signal processing strategies. This may be of particular use for haptic stimulation in combination with an additional beamforming module added to the proposed haptic signal processing strategies. This may allow for better speech-in-noise performance than was demonstrated by Fletcher et al. (2020b) for speech in spatially separated noise. This may also offer additional sound localisation performance enhancements over those demonstrated by Fletcher et al. (2020a). Significant adaption of the presented methods and use of head-mounted electrodes (potentially behind the ears) would be needed for this to provide an effective real-world solution. As with other interventions, the TRF may also be used for objective, automatic clinical fitting of a haptic intervention. As with other interventions, this would require significant reduction in analysis time and require fewer electrodes to be clinically feasible. With solutions to these issues the TRF may allow for fitting of haptic interventions to infants, potentially maximising the performance benefits, similar to the observed benefits of early use of acoustic and cochlear implant intervention (Flexer, 2011). Exploration of this area is also timely at the given TRL 3–4 of current haptic device development, as it may provide considerable benefits to the performance of the device prior to its translation to clinical applications.

7.4 Conclusion

The aim of this thesis was to contribute to the improvement of speech-in-noise performance for the hearing impaired. Progress has been made, through development of two methods for enhanced assessment and improvement of speech-in-noise performance.

An existing neuroimaging method (the TRF) was adapted, use of an ElasticNet regression model and a perceptually motivated loudness feature, to improve its reconstruction performance. Results suggested that the proposed method provided improved reconstruction performance in comparison to previously proposed TRF implementations. Therefore it may be possible to adapt the proposed method for the improvement of existing TRF based speech-in-noise prediction methods. This would form the initial steps towards a clinically viable, automatic and objective speech-in-noise assessment method. This new method may additionally be a valuable new tool for analysis of the underlying neural mechanisms of speech-in-noise performance.

Two haptic neuroprosthetics were also developed. The first aimed to improve pitch discrimination performance for cochlear implant users. Evidence was presented, suggesting marked improvement in pitch discrimination performance. This improvement in discrimination of pitch, a key cue for speech and music perception, suggested that this method may provide benefit in these areas. The second neuroprosthetic aimed to improve speech-in-noise performance for cochlear implant users. Results suggested that this may provide benefit to speech-in-noise performance for certain users, but that the proposed design and accompanying training program may not be optimal. Recommendations were outlined for the improvement of the method in further developments towards a real-world usable haptic intervention for cochlear implant users.

In addition to the use of the presented neuroimaging and haptics methods individually, this thesis outlined the potential for the combination of these methods. This presents the exciting possibility that the TRF method could provide further insight into the underlying neural mechanisms that drive haptic improvement to speech-in-noise performance.

Appendix A

Individual participant results for Chapter 3

Further results from Chapter 2. Plots illustrate 3 unplanned post-hoc wilcoxons sign-ranked tests for each participant individually. Tests were performed on the 5 cross-validation folds of the best model generated for each participant. No correction for multiple comparisons was applied.

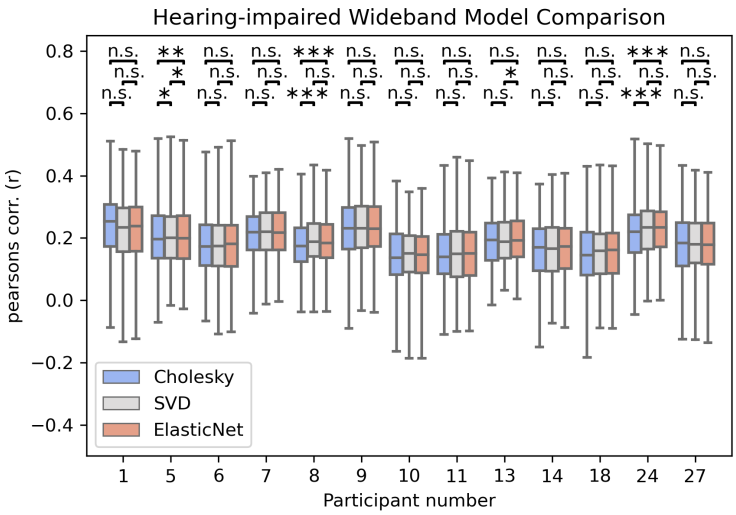

Figure A.1: Individual hearing-impaired subject correlation for the wideband Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$* * * : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

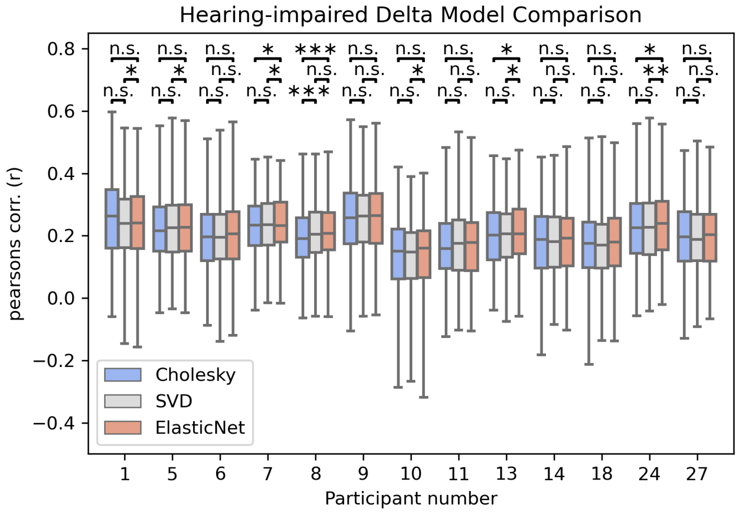

Figure A.2: Individual hearing-impaired subject correlation for the delta Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$* * * : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

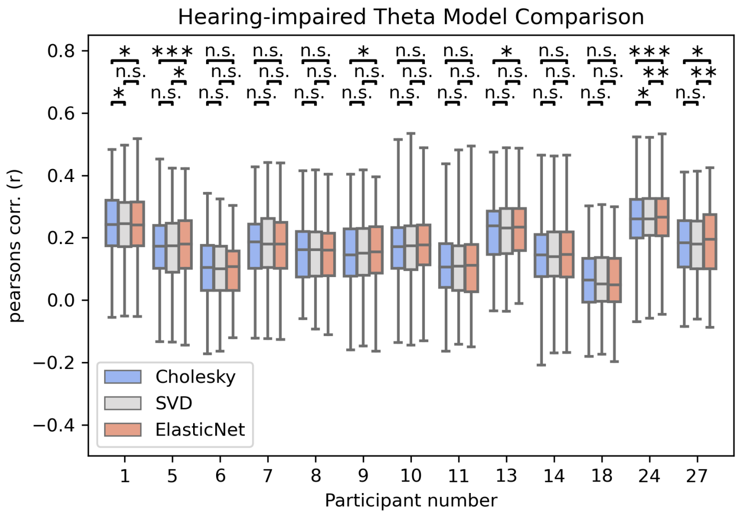

Figure A.3: Individual hearing-impaired subject correlation for the theta Cholesky Ridge, SVD Ridge and ElasticNet models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$* * * : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

Appendix B

Individual participant results for Chapter 4

Further results from Chapter 3. Figures B.1 to B.3 illustrate the 3 unplanned posthoc Wilcoxons sign-ranked tests for each hearing impaired participant individually. Tests were performed on the 5 cross-validation folds of the best model generated for each participant. No correction for multiple comparisons was applied. For the wideband and delta band conditions, all participants reconstructions were found to be significantly better than their time-reversed models. However, for the theta band model participant participant 18 ($W = 614, p = 0.053$) did not reach significance for the amplitude envelope model. For the loudness model participant 11($W = 774, p = 0.299$) was also not significant.

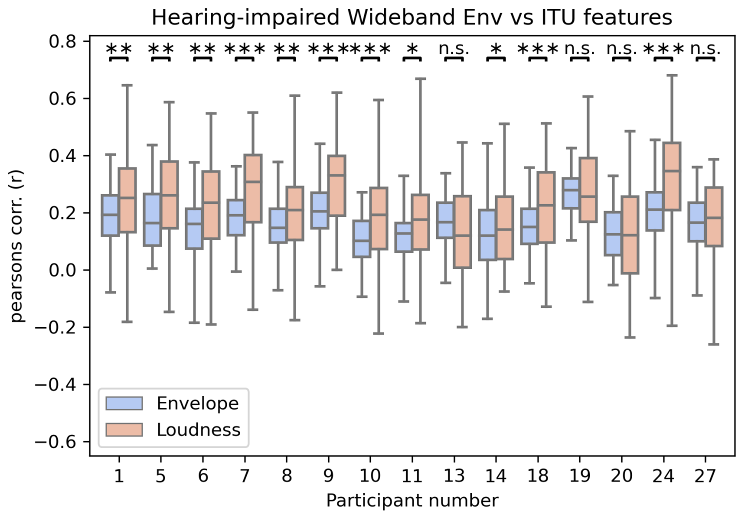

Figure B.1: Individual hearing-impaired subject correlation for the wideband amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $*** : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

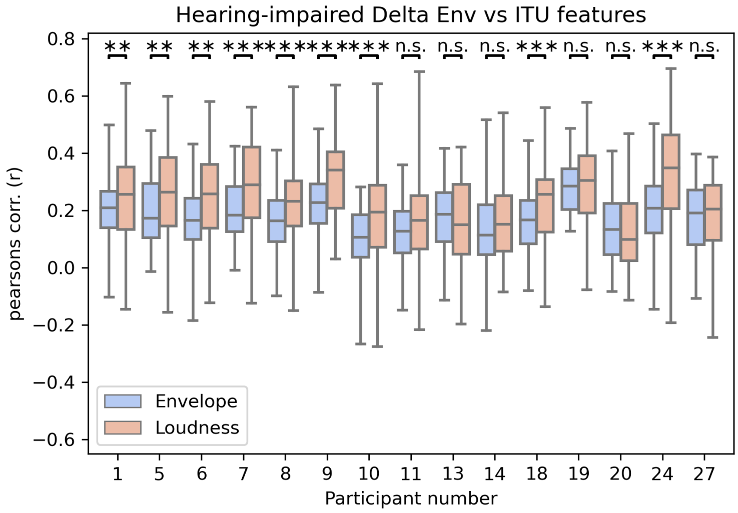

Figure B.2: Individual hearing-impaired subject correlation for the delta amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $*** : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

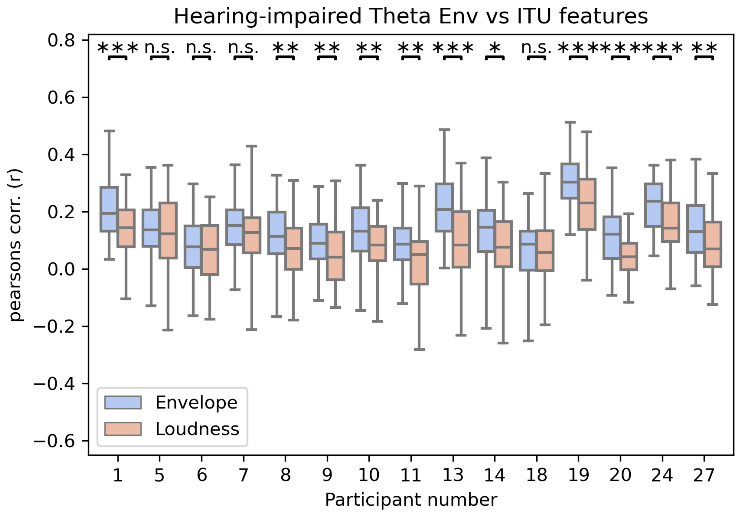

Figure B.3: Individual hearing-impaired subject correlation for the theta amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $*** : p \leqslant 0.001, ** : p \leqslant 0.01, * : p \leqslant 0.05, \text{n.s.} : p > 0.05$

Figures B.1 to B.3 illustrate the 3 unplanned post-hoc wilcoxons sign-ranked tests for each normal hearing participant individually. For the wideband envelope model, participant 5's model did not perform significantly better than the time-reversed model ($W = 655.0, p = 0.056$). For the wideband loudness model, participants 24 ($W = 620.0, p = 0.060$) and 28 ($W = 634.0, p = 0.077$) also did not reach significance. For the delta band envelope model, participant 5's model again did not reach significance ($W = 708.0, p = 0.128$). Also, as with the wideband loudness model, the delta loudness model for participants 24 ($W = 622.0, p = 0.062$) and 28 ($W = 649.0, p = 0.100$) were not significant. For the theta band envelope, models that were not significantly better than their time reversed equivalents are detailed in Table B.1. For the loudness model, results are presented in Table B.2.

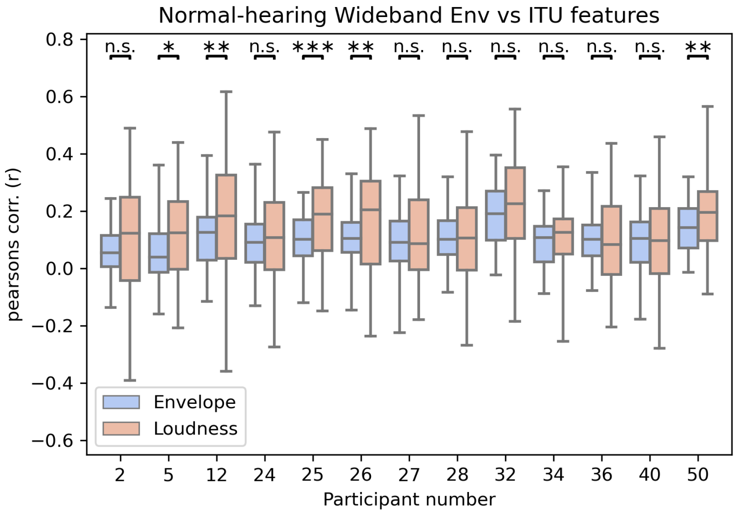

Figure B.4: Individual normal-hearing subject correlation for the wideband amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$***: p \leqslant 0.001, **: p \leqslant 0.01, *: p \leqslant 0.05, \text{n.s.}: p > 0.05$

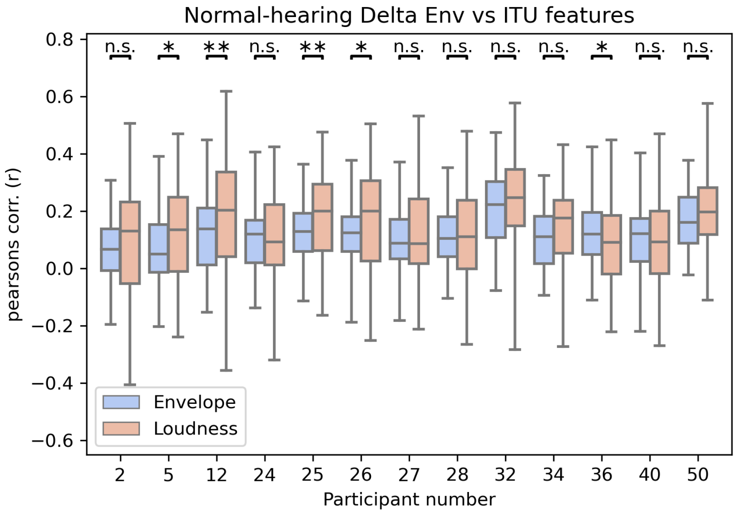

Figure B.5: Individual normal-hearing subject correlation for the delta amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as:
$***: p \leqslant 0.001, **: p \leqslant 0.01, *: p \leqslant 0.05, \text{n.s.}: p > 0.05$

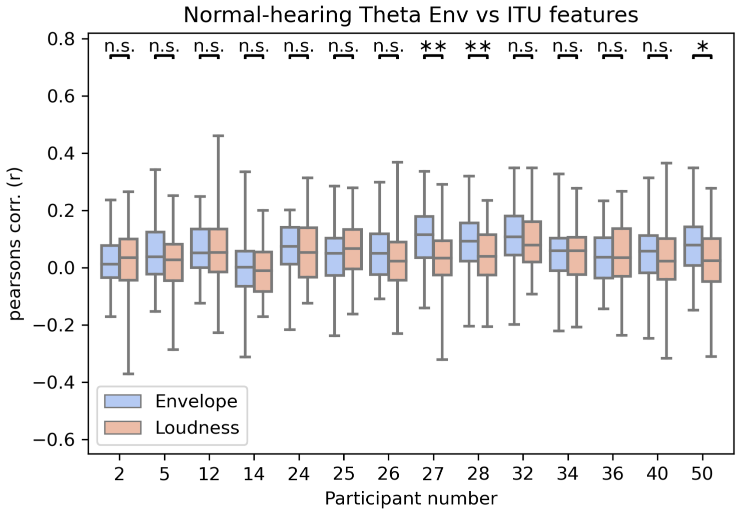

Figure B.6: Individual hearing-impaired subject correlation for the theta amplitude envelope and perceptual loudness feature reconstruction models. Differences in distributions (assessed using Wilcoxons sign-ranked tests) are annotated as: $*\!*\!*: p \leqslant 0.001, **: p \leqslant 0.01, *: p \leqslant 0.05, \text{n.s.}: p > 0.05$

Participant no.	W	p
P02	838.0	1.000
P12	747.0	0.432
P14	759.0	0.752
P25	765.0	0.472
P40	609.0	0.073

Table B.1: Wilcoxon test results for envelope feature models that did not reach significance in the theta band

Participant no.	W	p
P02	639.0	0.127
P05	871.0	0.746
P12	641.0	0.131
P14	839.0	0.950
P26	659.0	0.119
P27	745.0	0.211
P28	800.0	0.397
P34	808.0	0.862
P36	648.0	0.099
P40	862.0	1.000

Table B.2: Wilcoxon test results for loudness feature models that did not reach significance in the theta band

References

Abberton, E. and A. J. Fourcin (1978). “Intonation and Speaker Identification”. In: Language and Speech 21.4, pp. 305–318. issn: 0023-8309. doi: 10.1177/002383097802100405. pmid: 750790.

Accou, B. et al. (2021). “Modeling the Relationship between Acoustic Stimulus and EEG with a Dilated Convolutional Neural Network”. In: 2020 28th European Signal Processing Conference (EUSIPCO). 2020 28th European Signal Processing Conference (EUSIPCO), pp. 1175–1179. doi: 10.23919/Eusipco47968.2020.9287417.

Adank, P. et al. (2009). “Comprehension of Familiar and Unfamiliar Native Accents under Adverse Listening Conditions.” In: Journal of Experimental Psychology: Human Perception and Performance 35.2, pp. 520–529. issn: 1939-1277, 0096-1523. doi: 10.1037/a0013552. url: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0013552 (visited on 12/02/2019).

Aiken, S. J. and T. W. Picton (2008). “Human Cortical Responses to the Speech Envelope:” in: Ear and Hearing 29.2, pp. 139–157. issn: 0196-0202. doi: 10.1097/AUD.0b013e31816453dc. url: https://insights.ovid.com/crossref?an=00003446-200804000-00001 (visited on 12/02/2019).

Aitkin, L. M., C. E. Kenyon, and P. Philpott (1981). “The Representation of the Auditory and Somatosensory Systems in the External Nucleus of the Cat Inferior Colliculus”. In: The Journal of Comparative Neurology 196.1, pp. 25–40. issn: 0021-9967. doi: 10.1002/cne.901960104. pmid: 7204665.

Akbari, H. et al. (2019). “Towards Reconstructing Intelligible Speech from the Human Auditory Cortex”. In: Scientific Reports 9.1 (1), p. 874. issn: 2045-2322. doi: 10.1038/s41598-018-37359-z. url: https://www.nature.com/articles/s41598-018-37359-z (visited on 10/30/2021).

Akinpelu, O. V. et al. (2014). “Otoacoustic Emissions in Newborn Hearing Screening: A Systematic Review of the Effects of Different Protocols on Test Outcomes”. In: International Journal of Pediatric Otorhinolaryngology 78.5, pp. 711–717. issn: 0165-5876. doi: 10.1016/j.ijporl.2014.01.021. url: https://www.sciencedirect.com/science/article/pii/S0165587614000500 (visited on 04/17/2022).

Alexander, J. (2013). “Individual Variability in Recognition of Frequency-Lowered Speech”. In: Seminars in Hearing 34.02, pp. 086–109. issn: 0734-0451, 1098-8955. doi: 10.1055/

s-0033-1341346. url: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0033-1341346 (visited on 09/05/2021).

Anderson, S. et al. (2010a). "Neural Timing Is Linked to Speech Perception in Noise". In: Journal of Neuroscience 30.14, pp. 4922–4926. issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0107-10.2010. url: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0107-10.2010 (visited on 12/02/2019).

Anderson, S. and N. Kraus (2010). "Objective Neural Indices of Speech-in-Noise Perception". In: Trends in Amplification 14.2, pp. 73–83. issn: 1084-7138. doi: 10.1177/1084713810380227. url: https://doi.org/10.1177/1084713810380227 (visited on 07/27/2021).

Anderson, S. et al. (2010b). "Brainstem Correlates of Speech-in-Noise Perception in Children". In: Hearing Research 270.1-2, pp. 151–157. issn: 03785955. doi: 10.1016/j.heares.2010.08.001. url: https://linkinghub.elsevier.com/retrieve/pii/S0378595510003552 (visited on 12/02/2019).

Anderson, S. et al. (2013). "Auditory Brainstem Response to Complex Sounds Predicts Self-Reported Speech-in-Noise Performance". In: Journal of Speech, Language, and Hearing Research 56.1, pp. 31–43. issn: 1092-4388, 1558-9102. doi: 10.1044/1092-4388(2012/12-0043). url: http://pubs.asha.org/doi/10.1044/1092-4388%282012/12-0043%29 (visited on 12/02/2019).

Anderson-Hsieh, J. and K. Koehler (1988). "The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension*". In: Language Learning 38.4, pp. 561–613. issn: 00238333. doi: 10.1111/j.1467-1770.1988.tb00167.x. url: http://doi.wiley.com/10.1111/j.1467-1770.1988.tb00167.x (visited on 12/02/2019).

Arduino (2021). Arduino Due. Version 1.0. Somerville, MA: Arduino. url: https://store.arduino.cc/products/arduino-due?selectedStore=eu (visited on 10/22/2021).

Aroudi, A. and S. Doclo (2020). "Cognitive-Driven Binaural Beamforming Using EEG-Based Auditory Attention Decoding". In: IEEE/ACM Transactions on Audio, Speech, and Language Processing 28, pp. 862–875. issn: 2329-9304. doi: 10.1109/TASLP.2020.2969779.

Bach-y-Rita, P. (2004). "Tactile Sensory Substitution Studies". In: Annals of the New York Academy of Sciences 1013, pp. 83–91. issn: 0077-8923. doi: 10.1196/annals.1305.006. pmid: 15194608.

Banse, R. and K. R. Scherer (1996). "Acoustic Profiles in Vocal Emotion Expression". In: Journal of Personality and Social Psychology 70.3, pp. 614–636. issn: 0022-3514. doi: 10.1037//0022-3514.70.3.614. pmid: 8851745.

Barrenas, M.-L. and I. Wikstrom (2000). "The Influence of Hearing and Age on Speech Recognition Scores in Noise in Audiological Patients and in the General Population:" in: Ear and Hearing 21.6, pp. 569–577. issn: 0196-0202. doi: 10.1097/00003446-200012000-00004. url: https://insights.ovid.com/crossref?an=00003446-200012000-00004 (visited on 12/02/2019).

Bartkiw, B. (1988). “Reducing the Stigma of Deafness-Hearing Aids with Enhanced Visual Appeal”. In: British Journal of Audiology 22.3, pp. 167–169. issn: 0300-5364. doi: 10.3109/03005368809076448. pmid: 3167254. url: https://doi.org/10.3109/03005368809076448 (visited on 09/02/2021).

Basura, G. J., S. D. Koehler, and S. E. Shore (2012). “Multi-Sensory Integration in Brainstem and Auditory Cortex”. In: Brain Research 1485, pp. 95–107. issn: 1872-6240. doi: 10.1016/j.brainres.2012.08.037. pmid: 22995545.

Bear, M. F., B. W. Connors, and M. A. Paradiso (2016). Neuroscience: Exploring the Brain. Fourth edition. Philadelphia: Wolters Kluwer. 975 pp. isbn: 978-0-7817-7817-6.

Bednar, A. and E. C. Lalor (2020). “Where Is the Cocktail Party? Decoding Locations of Attended and Unattended Moving Sound Sources Using EEG”. In: NeuroImage 205, p. 116283. issn: 1095-9572. doi: 10.1016/j.neuroimage.2019.116283. pmid: 31629828.

Bench, J., A. Kowal, and J. Bamford (1979). “The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children”. In: British Journal of Audiology 13.3, pp. 108–112. issn: 0300-5364. doi: 10.3109/03005367909078884. pmid: 486816.

Bentler, R. and L.-K. Chiou (2006). “Digital Noise Reduction: An Overview”. In: Trends in Amplification 10.2, pp. 67–82. issn: 1084-7138. doi: 10.1177/1084713806289514. url: https://doi.org/10.1177/1084713806289514 (visited on 09/05/2021).

Bentler, R. et al. (2008). “Digital Noise Reduction: Outcomes from Laboratory and Field Studies”. In: International Journal of Audiology 47.8, pp. 447–460. issn: 1708-8186. doi: 10.1080/14992020802033091. pmid: 18698521.

Bernstein, L. E. et al. (2008). “Spatiotemporal Dynamics of Audiovisual Speech Processing”. In: NeuroImage 39.1, pp. 423–435. issn: 1053-8119. doi: 10.1016/j.neuroimage.2007.08.035. url: https://www.sciencedirect.com/science/article/pii/S1053811907007689 (visited on 07/28/2021).

Bidelman, G. M. (2016). “Relative Contribution of Envelope and Fine Structure to the Subcortical Encoding of Noise-Degraded Speech”. In: The Journal of the Acoustical Society of America 140.4, EL358–EL363. issn: 0001-4966. doi: 10.1121/1.4965248. url: http://asa.scitation.org/doi/10.1121/1.4965248 (visited on 12/02/2019).

Biesmans, W. et al. (2017). “Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario”. In: IEEE Transactions on Neural Systems and Rehabilitation Engineering 25.5, pp. 402–412. issn: 1534-4320, 1558-0210. doi: 10.1109/TNSRE.2016.2571900. url: https://ieeexplore.ieee.org/document/7478117/ (visited on 10/29/2021).

Biggs, J. et al. (2021). “A Natively Flexible 32-Bit Arm Microprocessor”. In: Nature 595.7868 (7868), pp. 532–536. issn: 1476-4687. doi: 10.1038/s41586-021-03625-w. url: https://www.nature.com/articles/s41586-021-03625-w (visited on 07/23/2021).

Billings, C. J. et al. (2013). “Predicting Perception in Noise Using Cortical Auditory Evoked Potentials”. In: Journal of the Association for Research in Otolaryngology

14.6, pp. 891–903. issn: 1525-3961, 1438-7573. doi: 10.1007/s10162-013-0415-y. url: http://link.springer.com/10.1007/s10162-013-0415-y (visited on 12/02/2019).

Bin Afif, A. et al. (2019). "Using Vibrotactile Stimulation to Improve Speech-in-Noise Performance for Cochlear Implant Users". In: British Society of Audiology Basic Auditory Science. University College London. url: https://www.ucl.ac.uk/ear/sites/ear/files/bas2019_programme.pdf.

Binder, J. (2000). "Human Temporal Lobe Activation by Speech and Nonspeech Sounds". In: Cerebral Cortex 10.5, pp. 512–528. issn: 14602199. doi: 10.1093/cercor/10.5.512. url: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/10.5.512 (visited on 12/02/2019).

BioSemi (n.d.). BioSemi ActiveTwo 32 Channel EEG System. Amsterdam: BioSemi. url: https://www.biosemi.com/products.htm.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Information Science and Statistics. New York: Springer. 738 pp. isbn: 978-0-387-31073-2.

Bispo, R. and V. Branco (2008). "Designing out Stigma : The Role of Objects in the Construction of Disabled People's Identity". In: British Journal of Audiology. url: https://www.semanticscholar.org/paper/Designing-out-stigma-%3A-the-role-of-objects-in-the-Bispo-Branco/1d7c049e8342e28aef0bcc692f5060ca6ddf02d1 (visited on 09/02/2021).

Blamey, P. J. and G. M. Clark (1985). "A Wearable Multiple-Electrode Electrotactile Speech Processor for the Profoundly Deaf". In: The Journal of the Acoustical Society of America 77.4, pp. 1619–1620. issn: 0001-4966. doi: 10.1121/1.392009. pmid: 3157717.

Boldt, R. et al. (2014). "Two-Point Tactile Discrimination Ability Is Influenced by Temporal Features of Stimulation". In: Experimental Brain Research 232.7, pp. 2179–2185. issn: 0014-4819, 1432-1106. doi: 10.1007/s00221-014-3908-y. url: http://link.springer.com/10.1007/s00221-014-3908-y (visited on 09/21/2021).

Boreas Technologies (2021). BOS1901 Piezo Haptic Driver with Digital Front End Datasheet. url: https://cdn.shopify.com/s/files/1/0066/6628/9221/files/BT001DDS01.02.pdf?v=1597953312 (visited on 09/23/2021).

Borg, E. (1982). "Correlation Between Auditory Brainstem Response (ABR) and Speech Discrimination Scores in Patients with Acoustic Neurinoma and in Patients with Cochlear Hearing Loss". In: Scandinavian Audiology 11.4, pp. 245–248. issn: 0105-0397. doi: 10.3109/01050398209087474. url: http://www.tandfonline.com/doi/full/10.3109/01050398209087474 (visited on 12/02/2019).

Bourguignon, M. et al. (2020). "Lip-Reading Enables the Brain to Synthesize Auditory Features of Unknown Silent Speech". In: Journal of Neuroscience 40.5, pp. 1053–1065. issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1101-19.2019. pmid: 31889007. url: https://www.jneurosci.org/content/40/5/1053 (visited on 08/24/2021).

Bregman, A. (1990). "Auditory Scene Analysis: The Perceptual Organization of Sound". In: Journal of The Acoustical Society of America - J ACOUST SOC AMER. Vol. 95. doi: 10.1121/1.408434.

British Society of Audiology (2018). Pure-Tone Air-Conduction and Bone- Conduction Threshold Audiometry with and without Masking. url: https://www.thebsa.org.uk/wp-content/uploads/2018/11/OD104-32-Recommended-Procedure-Pure-Tone-Audiometry-August-2018-FINAL.pdf (visited on 08/19/2021).

Brockmeier, S. J. et al. (2011). "The MuSIC Perception Test: A Novel Battery for Testing Music Perception of Cochlear Implant Users". In: Cochlear Implants International 12.1, pp. 10–20. issn: 1754-7628. doi: 10.1179/146701010X12677899497236. pmid: 21756454.

Brodbeck, C., A. Presacco, and J. Z. Simon (2018). "Neural Source Dynamics of Brain Responses to Continuous Stimuli: Speech Processing from Acoustics to Comprehension". In: NeuroImage 172, pp. 162–174. issn: 10538119. doi: 10.1016/j.neuroimage.2018.01.042. url: https://linkinghub.elsevier.com/retrieve/pii/S1053811918300429 (visited on 12/02/2019).

Brodbeck, C. and J. Z. Simon (2020). "Continuous Speech Processing". In: Current Opinion in Physiology 18, pp. 25–31. issn: 24688673. doi: 10.1016/j.cophys.2020.07.014. url: https://linkinghub.elsevier.com/retrieve/pii/S2468867320300766 (visited on 07/23/2021).

Bronkhorst, A. W. and R. Plomp (1988). "The Effect of Head-induced Interaural Time and Level Differences on Speech Intelligibility in Noise". In: The Journal of the Acoustical Society of America 83.4, pp. 1508–1516. issn: 0001-4966. doi: 10.1121/1.395906. url: https://asa.scitation.org/doi/10.1121/1.395906 (visited on 07/22/2021).

– (1992). "Effect of Multiple Speechlike Maskers on Binaural Speech Recognition in Normal and Impaired Hearing". In: The Journal of the Acoustical Society of America 92.6, pp. 3132–3139. issn: 0001-4966. doi: 10.1121/1.404209. url: http://asa.scitation.org/doi/10.1121/1.404209 (visited on 12/02/2019).

Bronkhorst, A. W. (2000). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions". In: 86, p. 12.

Brooks, P. L. et al. (1985). "Acquisition of a 250-word Vocabulary through a Tactile Vocoder". In: The Journal of the Acoustical Society of America 77.4, pp. 1576–1579. issn: 0001-4966. doi: 10.1121/1.392000. url: http://asa.scitation.org/doi/10.1121/1.392000 (visited on 09/24/2021).

Brown, C. A. and S. P. Bacon (2010). "Fundamental Frequency and Speech Intelligibility in Background Noise". In: Hearing research 266.1-2, pp. 52–59. issn: 0378-5955. doi: 10.1016/j.heares.2009.08.011. pmid: 19748564. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885573/ (visited on 07/23/2021).

Brüel & Kjær (2021a). Type 2250 G4 Sound Level Meter. Royston, Hertfordshire, UK.

Brüel & Kjær (2021b). Type 4157 Ear Simulator (IEC 60318-4 Coupler). Royston, Hertfordshire, UK. url: https://www.bksv.com/en/transducers/simulators/ear-mouth-simulators/4157.

– (2021c). Type 4231 Sound Calibrator. Royston, Hertfordshire, UK. url: https://www.bksv.com/en/transducers/simulators/ear-mouth-simulators/4157.

Brungart, D. S. (2001). "Informational and Energetic Masking Effects in the Perception of Two Simultaneous Talkers". In: The Journal of the Acoustical Society of America 109.3, pp. 1101–1109. issn: 0001-4966. doi: 10.1121/1.1345696. url: http://scitation.aip.org/content/asa/journal/jasa/109/3/10.1121/1.1345696 (visited on 12/02/2019).

Brungart, D. S. et al. (2006). "Isolating the Energetic Component of Speech-on-Speech Masking with Ideal Time-Frequency Segregation". In: The Journal of the Acoustical Society of America 120.6, pp. 4007–4018. issn: 0001-4966. doi: 10.1121/1.2363929. url: http://asa.scitation.org/doi/10.1121/1.2363929 (visited on 12/02/2019).

Bruns, L., D. Mürbe, and A. Hahne (2016). "Understanding Music with Cochlear Implants". In: Scientific Reports 6, p. 32026. issn: 2045-2322. doi: 10.1038/srep32026. pmid: 27558546. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4997320/ (visited on 09/28/2021).

BS.1770-2, I.-R. (2015). "Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level". In: ITU-R BS.1770-2 10, p. 25.

Budinger, E., P. Heil, and H. Scheich (2000). "Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones Unguiculatus). IV. Connections with Anatomically Characterized Subcortical Structures". In: The European Journal of Neuroscience 12.7, pp. 2452–2474. issn: 0953-816X. doi: 10.1046/j.1460-9568.2000.00143.x. pmid: 10947822.

Burguetti, F. A. R. and R. M. M. Carvallo (2008). "Efferent Auditory System: Its Effect on Auditory Processing". In: Brazilian Journal of Otorhinolaryngology 74.5, pp. 737–745. issn: 18088694. doi: 10.1016/S1808-8694(15)31385-9. url: https://linkinghub.elsevier.com/retrieve/pii/S1808869415313859 (visited on 12/07/2021).

Burkard, R. and K. Hecox (1983a). "The Effect of Broadband Noise on the Human Brainstem Auditory Evoked Response. I. Rate and Intensity Effects". In: The Journal of the Acoustical Society of America 74.4, pp. 1204–1213. issn: 0001-4966. doi: 10.1121/1.390024. url: http://asa.scitation.org/doi/10.1121/1.390024 (visited on 12/02/2019).

– (1983b). "The Effect of Broadband Noise on the Human Brainstem Auditory Evoked Response. II. Frequency Specificity". In: The Journal of the Acoustical Society of America 74.4, pp. 1214–1223. issn: 0001-4966. doi: 10.1121/1.390025. url: http://asa.scitation.org/doi/10.1121/1.390025 (visited on 12/02/2019).

Burr, D. et al. (2009). "Temporal Mechanisms of Multimodal Binding". In: Proceedings: Biological Sciences 276.1663, pp. 1761–1769. JSTOR: 30244009.

Buss, E. et al. (2009). "Masking Release for Words in Amplitude-Modulated Noise as a Function of Modulation Rate and Task". In: The Journal of the Acoustical Society of America 126.1, pp. 269–280. issn: 0001-4966. doi: 10.1121/1.3129506. url: http://asa.scitation.org/doi/10.1121/1.3129506 (visited on 12/02/2019).

Byrne, D. et al. (1994). "An International Comparison of Long-term Average Speech Spectra". In: The Journal of the Acoustical Society of America 96.4, pp. 2108–2120. issn: 0001-4966. doi: 10.1121/1.410152. url: https://asa.scitation.org/doi/10.1121/1.410152 (visited on 12/12/2021).

Calvert, G. A. et al. (1997). "Activation of Auditory Cortex during Silent Lipreading". In: Science (New York, N.Y.) 276.5312, pp. 593–596. issn: 0036-8075. doi: 10.1126/science.276.5312.593. pmid: 9110978.

Campbell, R. et al. (2001). "Cortical Substrates for the Perception of Face Actions: An fMRI Study of the Specificity of Activation for Seen Speech and for Meaningless Lower-Face Acts (Gurning)". In: Cognitive Brain Research 12.2, pp. 233–243. issn: 0926-6410. doi: 10.1016/S0926-6410(01)00054-4. url: https://www.sciencedirect.com/science/article/pii/S0926641001000544 (visited on 07/28/2021).

Carhart, R. and T. W. Tillman (1970). "Interaction of Competing Speech Signals With Hearing Losses". In: Archives of Otolaryngology - Head and Neck Surgery 91.3, pp. 273–279. issn: 0886-4470. doi: 10.1001/archotol.1970.00770040379010. url: http://archotol.jamanetwork.com/article.aspx?articleid=602843 (visited on 12/02/2019).

Carlyon, R. and T. Goehring (2021). "Cochlear Implant Research and Development in the Twenty-first Century: A Critical Update". In: Journal of the Association for Research in Otolaryngology. doi: 10.1007/s10162-021-00811-5.

Chen, F., Y. Hu, and M. Yuan (2015). "Evaluation of Noise Reduction Methods for Sentence Recognition by Mandarin-speaking Cochlear Implant Listeners". In: Ear and Hearing 36.1, pp. 61–71. issn: 1538-4667. doi: 10.1097/AUD.0000000000000074. pmid: 25127321.

Chen, J. et al. (2021). "Effects of Wireless Remote Microphone on Speech Recognition in Noise for Hearing Aid Users in China". In: Frontiers in Neuroscience 15, p. 389. issn: 1662-453X. doi: 10.3389/fnins.2021.643205. url: https://www.frontiersin.org/article/10.3389/fnins.2021.643205 (visited on 09/04/2021).

Cherry, E. C. (1953). "Some Experiments on the Recognition of Speech, with One and with Two Ears". In: The Journal of the Acoustical Society of America 25.5, pp. 975–979. issn: 0001-4966. doi: 10.1121/1.1907229. url: http://asa.scitation.org/doi/10.1121/1.1907229 (visited on 12/02/2019).

Chesnaye, M. A. et al. (2019). "A Group Sequential Test for ABR Detection". In: International Journal of Audiology 58.10, pp. 618–627. issn: 1499-2027. doi: 10.1080/14992027.2019.1625486. pmid: 31259611. url: https://doi.org/10.1080/14992027.2019.1625486 (visited on 10/30/2021).

Cholewiak, R. W. and A. A. Collins (2003). “Vibrotactile Localization on the Arm: Effects of Place, Space, and Age”. In: Perception & Psychophysics 65.7, pp. 1058–1077. issn: 1532-5962. doi: 10.3758/BF03194834. url: https://doi.org/10.3758/BF03194834 (visited on 09/21/2021).

Chung, K. and N. McKibben (2011). “Microphone Directionality, Pre-Emphasis Filter, and Wind Noise in Cochlear Implants”. In: Journal of the American Academy of Audiology 22.9, pp. 586–600. issn: 1050-0545. doi: 10.3766/jaaa.22.9.4. pmid: 22192604.

Chung, K. and F.-G. Zeng (2009). “Using Hearing Aid Adaptive Directional Microphones to Enhance Cochlear Implant Performance”. In: Hearing Research 250.1-2, pp. 27–37. issn: 1878-5891. doi: 10.1016/j.heares.2009.01.005. pmid: 19450437.

Chung, K., F.-G. Zeng, and K. N. Acker (2006). “Effects of Directional Microphone and Adaptive Multichannel Noise Reduction Algorithm on Cochlear Implant Performance”. In: The Journal of the Acoustical Society of America 120.4, pp. 2216–2227. issn: 0001-4966. doi: 10.1121/1.2258500. pmid: 17069317.

Clements, G. N. (1985). “The Geometry of Phonological Features”. In: Phonology 2.1, pp. 225–252. issn: 2059-6286, 0265-8062. doi: 10.1017/S0952675700000440. url: https://www.cambridge.org/core/journals/phonology/article/geometry-of-phonological-features/B9DC449706DDB5ACCA43AA885929B22C (visited on 07/23/2021).

Colum, P. (2021). LibriVox. LibriVox Audiobook. url: https://librivox.org/the-children-of-odin-by-padraic-colum/ (visited on 10/22/2021).

Coomes, D. L. and B. R. Schofield (2004). “Projections from the Auditory Cortex to the Superior Olivary Complex in Guinea Pigs”. In: The European Journal of Neuroscience 19.8, pp. 2188–2200. issn: 0953-816X. doi: 10.1111/j.0953-816X.2004.03317.x. pmid: 15090045.

Craig, J. C. (1972). “Difference Threshold for Intensity of Tactile Stimuli”. In: Perception & Psychophysics 11.2, pp. 150–152. issn: 1532-5962. doi: 10.3758/BF03210362. url: https://doi.org/10.3758/BF03210362 (visited on 09/21/2021).

Craig, J. C. and K. O. Johnson (2000). “The Two-Point Threshold: Not a Measure of Tactile Spatial Resolution”. In: Current Directions in Psychological Science 9.1, pp. 29–32. issn: 0963-7214. doi: 10.1111/1467-8721.00054. url: https://doi.org/10.1111/1467-8721.00054 (visited on 09/21/2021).

Crosse, M. J. et al. (2016). “The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli”. In: Frontiers in Human Neuroscience 10. issn: 1662-5161. doi: 10.3389/fnhum.2016.00604. url: http://journal.frontiersin.org/article/10.3389/fnhum.2016.00604/full (visited on 12/02/2019).

Cullington, H. E. and F.-G. Zeng (2008). “Speech Recognition with Varying Numbers and Types of Competing Talkers by Normal-Hearing, Cochlear-Implant, and Implant Simulation Subjects”. In: The Journal of the Acoustical Society of America 123.1, pp. 450–461. issn: 0001-4966. doi: 10.1121/1.2805617. url: http://asa.scitation.org/doi/10.1121/1.2805617 (visited on 12/02/2019).

Cycling 74 (2019). Max 8. Version 8.0.8. Walnut, CA, USA.

Darwin, C. (2008). “Listening to Speech in the Presence of Other Sounds”. In: Philosophical Transactions of the Royal Society B: Biological Sciences 363.1493, pp. 1011–1021. doi: 10.1098/rstb.2007.2156. url: https://royalsocietypublishing.org/doi/10.1098/rstb.2007.2156 (visited on 07/24/2021).

Das, N., A. Bertrand, and T. Francart (2018). “EEG-based Auditory Attention Detection: Boundary Conditions for Background Noise and Speaker Positions”. In: Journal of Neural Engineering 15.6, p. 066017. issn: 1741-2552. doi: 10.1088/1741-2552/aae0a6. url: https://doi.org/10.1088/1741-2552/aae0a6 (visited on 11/26/2021).

Daube, C., R. A. A. Ince, and J. Gross (2019). “Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech”. In: Current Biology 29.12, 1924–1937.e9. issn: 0960-9822. doi: 10.1016/j.cub.2019.04.067. url: https://www.sciencedirect.com/science/article/pii/S0960982219304968 (visited on 10/29/2021).

David, M. et al. (2017). “Sequential Stream Segregation of Voiced and Unvoiced Speech Sounds Based on Fundamental Frequency”. In: Hearing Research 344, pp. 235–243. issn: 0378-5955. doi: 10.1016/j.heares.2016.11.016. url: https://www.sciencedirect.com/science/article/pii/S037859551630315X (visited on 11/06/2021).

Dawson, P. W., S. J. Mauger, and A. A. Hersbach (–June 2011). “Clinical Evaluation of Signal-to-Noise Ratio–Based Noise Reduction in Nucleus® Cochlear Implant Recipients”. In: Ear and Hearing 32.3, pp. 382–390. issn: 0196-0202. doi: 10.1097/AUD.0b013e318201c200. url: https://journals.lww.com/ear-hearing/Abstract/2011/05000/Clinical_Evaluation_of_Signal_to_Noise_Ratio_Based.11.aspx (visited on 09/06/2021).

De Cheveigne, A. and H. Kawahara (2002). “YIN, a Fundamental Frequency Estimator for Speech and Musica)”. In: J. Acoust. Soc. Am. 111.4, p. 14.

Decruy, L., J. Vanthornhout, and T. Francart (2018). “Evidence for Enhanced Neural Tracking of the Speech Envelope Underlying Age-Related Speech-in-Noise Difficulties”. In: doi: 10.1101/489237. url: http://biorxiv.org/lookup/doi/10.1101/489237 (visited on 12/02/2019).

– (2020). “Hearing Impairment Is Associated with Enhanced Neural Tracking of the Speech Envelope”. In: Hearing Research 393, p. 107961. issn: 0378-5955. doi: 10.1016/j.heares.2020.107961. url: https://www.sciencedirect.com/science/article/pii/S0378595519304940 (visited on 11/25/2021).

De Taillez, T., B. Kollmeier, and B. T. Meyer (2020). “Machine Learning for Decoding Listeners’ Attention from Electroencephalography Evoked by Continuous Speech”. In: The European Journal of Neuroscience 51.5, pp. 1234–1241. issn: 1460-9568. doi: 10.1111/ejn.13790. pmid: 29205588.

Di Liberto, G. M. and E. C. Lalor (2017). "Indexing Cortical Entrainment to Natural Speech at the Phonemic Level: Methodological Considerations for Applied Research". In: Hearing Research 348, pp. 70–77. issn: 03785955. doi: 10.1016/j.heares.2017.02.015. url: https://linkinghub.elsevier.com/retrieve/pii/S0378595516304701 (visited on 12/02/2019).

Di Liberto, G. M., E. C. Lalor, and R. E. Millman (2018). "Causal Cortical Dynamics of a Predictive Enhancement of Speech Intelligibility". In: NeuroImage 166, pp. 247–258. issn: 10538119. doi: 10.1016/j.neuroimage.2017.10.066. url: https://linkinghub.elsevier.com/retrieve/pii/S1053811917309023 (visited on 12/02/2019).

Di Liberto, G. M., J. A. O'Sullivan, and E. C. Lalor (2015). "Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing". In: Current Biology 25.19, pp. 2457–2465. issn: 09609822. doi: 10.1016/j.cub.2015.08.030. url: https://linkinghub.elsevier.com/retrieve/pii/S0960982215010015 (visited on 12/02/2019).

Dimitrijevic, A., M. S. John, and T. W. Picton (2004). "Auditory Steady-State Responses and Word Recognition Scores in Normal-Hearing and Hearing-Impaired Adults:" in: Ear and Hearing 25.1, pp. 68–84. issn: 0196-0202. doi: 10.1097/01.AUD.0000111545.71693.48. url: https://insights.ovid.com/crossref?an=00003446-200402000-00007 (visited on 12/02/2019).

Dimitrijevic, A. et al. (2001). "Human Auditory Steady-State Responses to Tones Independently Modulated in Both Frequency and Amplitude:" in: Ear and Hearing 22.2, pp. 100–111. issn: 0196-0202. doi: 10.1097/00003446-200104000-00003. url: https://insights.ovid.com/crossref?an=00003446-200104000-00003 (visited on 12/02/2019).

Ding, N. and J. Z. Simon (2012). "Emergence of Neural Encoding of Auditory Objects While Listening to Competing Speakers". In: Proceedings of the National Academy of Sciences 109.29, pp. 11854–11859. issn: 0027-8424, 1091-6490. doi: 10.1073/pnas.1205381109. url: http://www.pnas.org/cgi/doi/10.1073/pnas.1205381109 (visited on 12/02/2019).

Ding, N. and J. Z. Simon (2014). "Cortical Entrainment to Continuous Speech: Functional Roles and Interpretations". In: Frontiers in Human Neuroscience 8. issn: 1662-5161. doi: 10.3389/fnhum.2014.00311. url: http://journal.frontiersin.org/article/10.3389/fnhum.2014.00311/abstract (visited on 12/02/2019).

Dobie, R. and S. Van Hemel (2004). Hearing Loss: Determining Eligibility for Social Security Benefits. Washington, D.C.: National Academies Press. isbn: 978-0-309-09296-8. doi: 10.17226/11099. url: http://www.nap.edu/catalog/11099 (visited on 12/02/2019).

Dorman, M. F. et al. (2016). "Sound Source Localization by Normal-Hearing Listeners, Hearing-Impaired Listeners and Cochlear Implant Listeners". In: Audiology and Neurotology 21.3, pp. 127–131. issn: 1420-3030, 1421-9700. doi: 10.1159/000444740.

pmid: 27077663. url: https://www.karger.com/Article/FullText/444740 (visited on 12/12/2021).

Dormann, C. F. et al. (2013). "Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance". In: Ecography 36.1, pp. 27–46. issn: 1600-0587. doi: 10.1111/j.1600-0587.2012.07348.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0587.2012.07348.x (visited on 10/23/2021).

Doucet, J. R., L. Rose, and D. K. Ryugo (2002). "The Cellular Origin of Corticofugal Projections to the Superior Olivary Complex in the Rat". In: Brain Research 925.1, pp. 28–41. issn: 0006-8993. doi: 10.1016/s0006-8993(01)03248-6. pmid: 11755898.

Drennan, W. R. et al. (2015). "Clinical Evaluation of Music Perception, Appraisal and Experience in Cochlear Implant Users". In: International Journal of Audiology 54.2, pp. 114–123. issn: 1708-8186. doi: 10.3109/14992027.2014.948219. pmid: 25177899.

Dreschler, W. A. et al. (2001). "ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment. International Collegium for Rehabilitative Audiology". In: Audiology: Official Organ of the International Society of Audiology 40.3, pp. 148–157. issn: 0020-6091. pmid: 11465297.

Drullman, R., J. M. Festen, and R. Plomp (1994). "Effect of Temporal Envelope Smearing on Speech Reception". In: The Journal of the Acoustical Society of America 95.2, pp. 1053–1064. issn: 0001-4966. doi: 10.1121/1.408467. pmid: 8132899.

Drullman, R. and A. W. Bronkhorst (2000). "Multichannel Speech Intelligibility and Talker Recognition Using Monaural, Binaural, and Three-Dimensional Auditory Presentation". In: The Journal of the Acoustical Society of America 107.4, pp. 2224–2235. issn: 0001-4966. doi: 10.1121/1.428503. url: http://asa.scitation.org/doi/10.1121/1.428503 (visited on 12/02/2019).

– (2004). "Speech Perception and Talker Segregation: Effects of Level, Pitch, and Tactile Support with Multiple Simultaneous Talkers". In: The Journal of the Acoustical Society of America 116.5, pp. 3090–3098. issn: 0001-4966. doi: 10.1121/1.1802535. url: http://asa.scitation.org/doi/10.1121/1.1802535 (visited on 12/02/2019).

Dubno, J. R., D. D. Dirks, and D. E. Morgan (1984). "Effects of Age and Mild Hearing Loss on Speech Recognition in Noise". In: The Journal of the Acoustical Society of America 76.1, pp. 87–96. issn: 0001-4966. doi: 10.1121/1.391011. url: http://asa.scitation.org/doi/10.1121/1.391011 (visited on 12/02/2019).

Dunn, C. C. et al. (2008). "Comparison of Speech Recognition and Localization Performance in Bilateral and Unilateral Cochlear Implant Users Matched on Duration of Deafness and Age at Implantation". In: Ear and Hearing 29.3, pp. 352–359. issn: 0196-0202. doi: 10.1097/AUD.0b013e318167b870. pmid: 18453885.

Dunn, C. C. et al. (2010). "Bilateral and Unilateral Cochlear Implant Users Compared on Speech Perception in Noise". In: Ear and hearing 31.2, pp. 296–298. issn: 0196-0202. doi: 10.1097/AUD.0b013e3181c12383. pmid: 19858720. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836420/ (visited on 09/06/2021).

Duquesnoy, A. J. (1983). “Effect of a Single Interfering Noise or Speech Source upon the Binaural Sentence Intelligibility of Aged Persons”. In: The Journal of the Acoustical Society of America 74.3, pp. 739–743. issn: 0001-4966. doi: 10.1121/1.389859. url: http://asa.scitation.org/doi/10.1121/1.389859 (visited on 12/02/2019).

Elliott, L. L. (1962). “Backward and Forward Masking of Probe Tones of Different Frequencies”. In: The Journal of the Acoustical Society of America 34.8, pp. 1116–1117. issn: 0001-4966. doi: 10.1121/1.1918254. url: https://asa.scitation.org/doi/abs/10.1121/1.1918254 (visited on 09/21/2021).

Erler, S. F. and D. C. Garstecki (2002). “Hearing Loss- and Hearing Aid-Related Stigma”. In: American Journal of Audiology 11.2, pp. 83–91. doi: 10.1044/1059-0889(2002/020). url: https://pubs.asha.org/doi/abs/10.1044/1059-0889%282002/020%29 (visited on 09/02/2021).

Ernestus, M., H. Baayen, and R. Schreuder (2002). “The Recognition of Reduced Word Forms”. In: Brain and Language 81.1-3, pp. 162–173. issn: 0093934X. doi: 10.1006/brln.2001.2514. url: https://linkinghub.elsevier.com/retrieve/pii/S0093934X01925143 (visited on 12/02/2019).

Ernst, M. O. and H. H. Bülthoff (2004). “Merging the Senses into a Robust Percept”. In: Trends in Cognitive Sciences 8.4, pp. 162–169. issn: 13646613. doi: 10.1016/j.tics.2004.02.002. url: https://linkinghub.elsevier.com/retrieve/pii/S1364661304000385 (visited on 08/15/2021).

Etard, O. and T. Reichenbach (2019). “Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise”. In: The Journal of Neuroscience 39.29, pp. 5750–5759. issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1828-18.2019. url: http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1828-18.2019 (visited on 12/02/2019).

Etymotics (2019). ER-2 Insert Earphones. IL, USA.

Fagan, M. K. (2015). “Cochlear Implantation at 12 Months: Limitations and Benefits for Vocabulary Production”. In: Cochlear Implants International 16.1, pp. 24–31. issn: 1467-0100. doi: 10.1179/1754762814Y.0000000075. pmid: 24954248. url: https://doi.org/10.1179/1754762814Y.0000000075 (visited on 04/16/2022).

Fay, R. R. et al., eds. (2016). Hearing Aids. 1st ed. 2016. Springer Handbook of Auditory Research 56. Preface – Introduction to Hearing Aids – Population of Hearing Aid Candidates – Hearing Aid Transducers – Speech Perception and Hearing Aids – Hearing Aid Signal Processing – Spatial Hearing and Interactions with Hearing Aids – Wireless Connectivity and Patient Interface – Fitting and Clinical Verification of Hearing Aid Performance – Validation of Hearing Aid Performance in Everyday Life – Listening to Music through Hearing Aids – Future Directions for Hearing Aid Development. Cham: Springer International Publishing : Imprint: Springer. 1 p. isbn: 978-3-319-33036-5. doi: 10.1007/978-3-319-33036-5.

Felmy, F. (2019). “The Nuclei of the Lateral Lemniscus”. In: The Oxford Handbook of the Auditory Brainstem. isbn: 978-0-19-084906-1. doi: 10.1093/oxfordhb/9780190849061.

013.13. url: https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190849061.001.0001/oxfordhb-9780190849061-e-13 (visited on 12/05/2021).

Feuerstein, J. F. (1992). “Monaural versus Binaural Hearing: Ease of Listening, Word Recognition, and Attentional Effort”. In: Ear and Hearing 13.2, p. 7.

Fletcher, M. (2021a). “Using Haptic Stimulation to Enhance Auditory Perception in Hearing-Impaired Listeners”. In: pmid: 33372550.

Fletcher, M. D. (2021b). “Can Haptic Stimulation Enhance Music Perception in Hearing-Impaired Listeners?” In: Frontiers in Neuroscience 15, p. 1123. issn: 1662-453X. doi: 10.3389/fnins.2021.723877. url: https://www.frontiersin.org/article/10.3389/fnins.2021.723877 (visited on 09/22/2021).

Fletcher, M. D., R. O. Cunningham, and S. R. Mills (2020a). “Electro-Haptic Enhancement of Spatial Hearing in Cochlear Implant Users”. In: Scientific Reports 10.1, p. 1621. issn: 2045-2322. doi: 10.1038/s41598-020-58503-8. url: http://www.nature.com/articles/s41598-020-58503-8 (visited on 07/29/2020).

Fletcher, M. D., S. R. Mills, and T. Goehring (2018). “Vibro-Tactile Enhancement of Speech Intelligibility in Multi-talker Noise for Simulated Cochlear Implant Listening”. In: Trends in Hearing 22. issn: 2331-2165, 2331-2165. doi: 10.1177/2331216518797838. url: http://journals.sagepub.com/doi/10.1177/2331216518797838 (visited on 12/26/2019).

Fletcher, M. D., H. Song, and S. W. Perry (2020b). “Electro-Haptic Stimulation Enhances Speech Recognition in Spatially Separated Noise for Cochlear Implant Users”. In: Scientific Reports 10.1. issn: 2045-2322. doi: 10.1038/s41598-020-69697-2. url: http://www.nature.com/articles/s41598-020-69697-2 (visited on 07/29/2020).

Fletcher, M. D., N. Thini, and S. W. Perry (2020c). “Enhanced Pitch Discrimination for Cochlear Implant Users with a New Haptic Neuroprosthetic”. In: Scientific Reports 10.1. issn: 2045-2322. doi: 10.1038/s41598-020-67140-0. url: http://www.nature.com/articles/s41598-020-67140-0 (visited on 07/29/2020).

Fletcher, M. D. and C. A. Verschuur (2021). “Electro-Haptic Stimulation: A New Approach for Improving Cochlear-Implant Listening”. In: Frontiers in Neuroscience 15, p. 613. issn: 1662-453X. doi: 10.3389/fnins.2021.581414. url: https://www.frontiersin.org/article/10.3389/fnins.2021.581414 (visited on 09/20/2021).

Fletcher, M. D., J. Zgheib, and S. W. Perry (2021a). “Sensitivity to Haptic Sound-Localisation Cues”. In: Scientific Reports 11.1, p. 312. issn: 2045-2322. doi: 10.1038/s41598-020-79150-z. url: http://www.nature.com/articles/s41598-020-79150-z (visited on 09/23/2021).

– (2021b). “Sensitivity to Haptic Sound-Localization Cues at Different Body Locations”. In: Sensors 21.11 (11), p. 3770. doi: 10.3390/s21113770. url: https://www.mdpi.com/1424-8220/21/11/3770 (visited on 07/20/2021).

Fletcher, M. D. et al. (2019). “Electro-Haptic Enhancement of Speech-in-Noise Performance in Cochlear Implant Users”. In: Scientific Reports 9.1, pp. 2045–2322. issn:

2045-2322. doi: 10.1038/s41598-019-47718-z. url: http://www.nature.com/articles/s41598-019-47718-z (visited on 12/26/2019).

Flexer, C. (2011). “Cochlear Implants and Neuroplasticity: Linking Auditory Exposure and Practice”. In: Cochlear Implants International 12 (sup1), S19–S21. issn: 1467-0100. doi: 10.1179/146701011X13001035752255. pmid: 21756466. url: https://doi.org/10.1179/146701011X13001035752255 (visited on 11/30/2021).

Florentine, M., S. Buus, and C. R. Mason (1987). “Level Discrimination as a Function of Level for Tones from 0.25 to 16 kHz”. In: The Journal of the Acoustical Society of America 81.5, pp. 1528–1541. issn: 0001-4966. doi: 10.1121/1.394505. pmid: 3584690.

Forte, A. E., O. Etard, and T. Reichenbach (2017). “The Human Auditory Brainstem Response to Running Speech Reveals a Subcortical Mechanism for Selective Attention”. In: p. 13.

Foxe, J. J. et al. (2000). “Multisensory Auditory-Somatosensory Interactions in Early Cortical Processing Revealed by High-Density Electrical Mapping”. In: Brain Research. Cognitive Brain Research 10.1-2, pp. 77–83. issn: 0926-6410. doi: 10.1016/s0926-6410(00)00024-0. pmid: 10978694.

Francart, T., A. Osses, and J. Wouters (2015). “Speech Perception with F0mod, a Cochlear Implant Pitch Coding Strategy”. In: International Journal of Audiology 54.6, pp. 424–432. issn: 1708-8186. doi: 10.3109/14992027.2014.989455. pmid: 25697275.

Friesen, L. M. et al. (2001). “Speech Recognition in Noise as a Function of the Number of Spectral Channels: Comparison of Acoustic Hearing and Cochlear Implants”. In: The Journal of the Acoustical Society of America 110.2, pp. 1150–1163. issn: 0001-4966. doi: 10.1121/1.1381538. pmid: 11519582.

Fu, Q. J., R. V. Shannon, and X. Wang (1998a). “Effects of Noise and Spectral Resolution on Vowel and Consonant Recognition: Acoustic and Electric Hearing”. In: The Journal of the Acoustical Society of America 104.6, pp. 3586–3596. issn: 0001-4966. doi: 10.1121/1.423941. pmid: 9857517.

Fu, Q.-J. and G. Nogaki (2005). “Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing”. In: JARO: Journal of the Association for Research in Otolaryngology 6.1, pp. 19–27. issn: 1525-3961. doi: 10.1007/s10162-004-5024-3. pmid: 15735937. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2504636/ (visited on 09/06/2021).

Fu, Q.-J. et al. (1998b). “Importance of Tonal Envelope Cues in Chinese Speech Recognition”. In: The Journal of the Acoustical Society of America 104.1, pp. 505–510. issn: 0001-4966. doi: 10.1121/1.423251. url: https://asa.scitation.org/doi/abs/10.1121/1.423251 (visited on 07/23/2021).

Fu, Z., X. Wu, and J. Chen (2019). “Congruent Audiovisual Speech Enhances Auditory Attention Decoding with EEG”. In: Journal of Neural Engineering 16.6, p. 066033. issn: 1741-2552. doi: 10.1088/1741-2552/ab4340. url: https://doi.org/10.1088/1741-2552/ab4340 (visited on 07/30/2021).

Fujisaki, W. and S. Nishida (2005). “Temporal Frequency Characteristics of Synchrony–Asynchrony Discrimination of Audio-Visual Signals”. In: Experimental Brain Research 166.3-4, pp. 455–464. issn: 0014-4819, 1432-1106. doi: 10.1007/s00221-005-2385-8. url: http://link.springer.com/10.1007/s00221-005-2385-8 (visited on 08/15/2021).

Galbraith, G. C. et al. (2004). “Brain Stem Evoked Response to Forward and Reversed Speech in Humans:” in: NeuroReport 15.13, pp. 2057–2060. issn: 0959-4965. doi: 10.1097/00001756-200409150-00012. url: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001756-200409150-00012 (visited on 12/02/2019).

Galvin, J. J. and Q.-J. Fu (2009). “Influence of Stimulation Rate and Loudness Growth on Modulation Detection and Intensity Discrimination in Cochlear Implant Users”. In: Hearing research 250.1-2, pp. 46–54. issn: 0378-5955. doi: 10.1016/j.heares.2009.01.009. pmid: 19450432. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5844469/ (visited on 09/21/2021).

Galvin, K. L. et al. (1999). “A Comparison of Tactaid II and Tactaid 7 Use by Adults with a Profound Hearing Impairment”. In: Ear and Hearing 20.6, pp. 471–482. issn: 0196-0202. url: https://journals.lww.com/ear-hearing/Abstract/1999/12000/A_Comparison_of_Tactaid_II_and_Tactaid_7_Use_by.3.aspx (visited on 08/12/2021).

Garadat, S. N. and B. E. Pfingst (2011). “Relationship between Gap Detection Thresholds and Loudness in Cochlear-Implant Users”. In: Hearing Research 275.1-2, pp. 130–138. issn: 1878-5891. doi: 10.1016/j.heares.2010.12.011. pmid: 21168479.

Geirnaert, S., T. Francart, and A. Bertrand (2020). “An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control”. In: IEEE Transactions on Neural Systems and Rehabilitation Engineering 28.1, pp. 307–317. issn: 1558-0210. doi: 10.1109/TNSRE.2019.2952724.

Geirnaert, S. et al. (2021). “EEG-based Auditory Attention Decoding: Towards Neuro-Steered Hearing Devices”. In: IEEE Signal Processing Magazine 38.4, pp. 89–102. issn: 1053-5888, 1558-0792. doi: 10.1109/MSP.2021.3075932. arXiv: 2008.04569. url: http://arxiv.org/abs/2008.04569 (visited on 10/30/2021).

Gelfand, S. A. (2016). Essentials of Audiology. 4th ed.

– (2018). Hearing: An Introduction to Psychological and Physiological Acoustics. Sixth edition. Physical concepts – Anatomy – Conductive mechanism – Cochlear mechanisms and processes – Auditory nerve – Auditory pathways – Psychoacoustic methods – Signal detection theory – Auditory sensitivity – Masking – Loudness – Pitch and timbre – Binaural and spatial hearing – Speech and its perception. Boca Raton: CRC Press. isbn: 978-1-4987-7542-7.

Gescheider, G. A. (1966). “Resolving of Successive Clicks by the Ears and Skin”. In: Journal of Experimental Psychology 71.3, pp. 378–381. issn: 0022-1015. doi: 10.1037/h0022950. pmid: 5908818.

Gescheider, G. A., S. J. Bolanowski, and R. T. Verrillo (1989). “Vibrotactile Masking: Effects of Stimulus Onset Asynchrony and Stimulus Frequency”. In: The Journal of the Acoustical Society of America 85.5, pp. 2059–2064. issn: 0001-4966. doi: 10.1121/1.397858. pmid: 2732386.

Gescheider, G. A., J. J. Zwislocki, and A. Rasmussen (1996). “Effects of Stimulus Duration on the Amplitude Difference Limen for Vibrotaction”. In: The Journal of the Acoustical Society of America 100 (4 Pt 1), pp. 2312–2319. issn: 0001-4966. doi: 10.1121/1.417940. pmid: 8865638.

Gescheider, G. A. (1967). “Auditory and Cutaneous Temporal Resolution of Successive Brief Stimuli”. In: Journal of Experimental Psychology 75.4, pp. 570–572. issn: 0022-1015. doi: 10.1037/h0025113.

Gescheider, G. A., J. H. Wright, and R. T. Verrillo (2010). Information-Processing Channels in the Tactile Sensory System. 0th ed. Psychology Press. isbn: 978-1-135-41925-7. doi: 10.4324/9780203890004. url: https://www.taylorfrancis.com/books/9781135419257 (visited on 07/21/2021).

Glasberg, B. R. and B. C. J. Moore (1986). “Auditory Filter Shapes in Subjects with Unilateral and Bilateral Cochlear Impairments”. In: The Journal of the Acoustical Society of America 79.4, pp. 1020–1033. issn: 0001-4966. doi: 10.1121/1.393374. url: http://asa.scitation.org/doi/10.1121/1.393374 (visited on 12/02/2019).

Goff, G. D. (1967). “Differential Discrimination of Frequency of Cutaneous Mechanical Vibration”. In: Journal of Experimental Psychology 74 (2, Pt.1), pp. 294–299. issn: 0022-1015. doi: 10.1037/h0024561.

Goldstein, E. B. (2011). Cognitive Psychology. Wadsworth Cengage Learning. 444 pp. isbn: 978-1-111-18588-6. Google Books: sr9VmgEACAAJ.

Gramfort, A. (2013). “MEG and EEG Data Analysis with MNE-Python”. In: Frontiers in Neuroscience 7. issn: 1662453X. doi: 10.3389/fnins.2013.00267. url: http://journal.frontiersin.org/article/10.3389/fnins.2013.00267/abstract (visited on 12/02/2019).

Grange, J. A. et al. (2017). “Cochlear Implant Simulator with Independent Representation of the Full Spiral Ganglion”. In: The Journal of the Acoustical Society of America 142.5, EL484–EL489. issn: 0001-4966. doi: 10.1121/1.5009602. url: http://asa.scitation.org/doi/10.1121/1.5009602 (visited on 12/02/2019).

Grant, K. W. and P. F. Seitz (2000). “The Recognition of Isolated Words and Words in Sentences: Individual Variability in the Use of Sentence Context”. In: p. 13.

Grant, K. W. and L. D. Braida (1991). “Evaluating the Articulation Index for Auditory–Visual Input”. In: The Journal of the Acoustical Society of America 89.6, pp. 2952–2960. issn: 0001-4966. doi: 10.1121/1.400733. url: https://asa.scitation.org/doi/abs/10.1121/1.400733 (visited on 08/15/2021).

Greenberg, S., ed. (2004). Speech Processing in the Auditory System. Springer Handbook of Auditory Research v. 18. New York: Springer. 476 pp. isbn: 978-0-387-00590-4.

Groenen, P. A. P. et al. (2001). “Speech-Evoked Cortical Potentials Recognition in Cochlear Implant Users and Speech”. In: Scandinavian Audiology 30.1, pp. 31–40. issn: 0105-0397. doi: 10.1080/010503901750069554. url: http://www.tandfonline.com/doi/full/10.1080/010503901750069554 (visited on 12/02/2019).

Guger, C., B. Z. Allison, and A. Gunduz, eds. (2021a). Brain-Computer Interface Research: A State-of-the-Art Summary 10. SpringerBriefs in Electrical and Computer Engineering. Cham: Springer International Publishing. isbn: 978-3-030-79286-2 978-3-030-79287-9. doi: 10.1007/978-3-030-79287-9. url: https://link.springer.com/10.1007/978-3-030-79287-9 (visited on 11/25/2021).

Guger, C., B. Z. Allison, and M. Tangermann (2021b). Brain-Computer Interface Research: A State-of-the-Art Summary 9. Springer Nature. 148 pp. isbn: 978-3-030-60460-8. Google Books: hBwnEAAAQBAJ.

Hagerman, B. (1982). “Sentences for Testing Speech Intelligibility in Noise”. In: Scandinavian Audiology 11.2, pp. 79–87. issn: 0105-0397. doi: 10.3109/01050398209076203. url: http://www.tandfonline.com/doi/full/10.3109/01050398209076203 (visited on 12/02/2019).

Hairston, W. D. et al. (2003). “Multisensory Enhancement of Localization under Conditions of Induced Myopia”. In: Experimental Brain Research 152.3, pp. 404–408. issn: 0014-4819, 1432-1106. doi: 10.1007/s00221-003-1646-7. url: http://link.springer.com/10.1007/s00221-003-1646-7 (visited on 08/15/2021).

Hall, J. W. (1992). Handbook of Auditory Evoked Responses. Boston: Allyn and Bacon. 871 pp. isbn: 978-0-205-13566-0.

Halle, M. and K. N. Stevens (1971). A Note on Laryngeal Features. De Gruyter Mouton, pp. 45–61. isbn: 978-3-11-087125-8. url: https://www.degruyter.com/document/doi/10.1515/9783110871258.45/html (visited on 07/23/2021).

Hammerschmidt, K. and U. Jürgens (2007). “Acoustical Correlates of Affective Prosody”. In: Journal of Voice 21.5, pp. 531–540. issn: 0892-1997. doi: 10.1016/j.jvoice.2006.03.002. url: https://www.sciencedirect.com/science/article/pii/S0892199706000361 (visited on 07/24/2021).

“Applications of Adaptive Signal Processing Methods in High-End Hearing Aids” (2006). In: Topics in Acoustic Echo and Noise Control: Selected Methods for the Cancellation of Acoustical Echoes, the Reduction of Background Noise, and Speech Processing. Ed. by E. Hänsler and G. Schmidt. Signals and Communication Technology. Berlin, Heidelberg: Springer, pp. 599–636. isbn: 978-3-540-33213-8. doi: 10.1007/3-540-33213-8_15. url: https://doi.org/10.1007/3-540-33213-8_15 (visited on 09/04/2021).

Harris, J. D. (1963). Loudness Discrimination. American Speech and Hearing Association].

Hawley, M. L., R. Y. Litovsky, and H. S. Colburn (1999). “Speech Intelligibility and Localization in a Multi-Source Environment”. In: The Journal of the Acoustical Society of America 105.6, pp. 3436–3448. issn: 0001-4966. doi: 10.1121/1.424670. url: http://asa.scitation.org/doi/10.1121/1.424670 (visited on 12/02/2019).

Herdman, A. T. et al. (2002). "Intracerebral Sources of Human Auditory Steady-State Responses". In: Brain Topography 15.2, pp. 69–86. issn: 08960267. doi: 10.1023/A:1021470822922. url: http://link.springer.com/10.1023/A:1021470822922 (visited on 09/03/2021).

Hess, W. (2012). Pitch Determination of Speech Signals: Algorithms and Devices. Springer Science & Business Media. 713 pp. isbn: 978-3-642-81926-1. Google Books: VfTOCAAAQBAJ.

Hickok, G. and D. Poeppel (2007). "The Cortical Organization of Speech Processing". In: Nature Reviews Neuroscience 8.5, pp. 393–402. issn: 1471-003X, 1471-0048. doi: 10.1038/nrn2113. url: http://www.nature.com/articles/nrn2113 (visited on 12/02/2019).

Hillenbrand, J. M., R. A. Houde, and R. T. Gayvert (2006). "Speech Perception Based on Spectral Peaks versus Spectral Shape". In: J. Acoust. Soc. Am. 119.6, p. 15.

Hochmair, I. et al. (2015). "Deep Electrode Insertion and Sound Coding in Cochlear Implants". In: Hearing Research. Lasker Award 322, pp. 14–23. issn: 0378-5955. doi: 10.1016/j.heares.2014.10.006. url: https://www.sciencedirect.com/science/article/pii/S0378595514001701 (visited on 09/07/2021).

Hoen, M. et al. (2007). "Phonetic and Lexical Interferences in Informational Masking during Speech-in-Speech Comprehension". In: Speech Communication 49.12, pp. 905–916. issn: 01676393. doi: 10.1016/j.specom.2007.05.008. url: https://linkinghub.elsevier.com/retrieve/pii/S0167639307000970 (visited on 12/02/2019).

Hoffmann, S. and M. Falkenstein (2008). "The Correction of Eye Blink Artefacts in the EEG: A Comparison of Two Prominent Methods". In: PLoS ONE 3.8. Ed. by T. Bussey, e3004. issn: 1932-6203. doi: 10.1371/journal.pone.0003004. url: https://dx.plos.org/10.1371/journal.pone.0003004 (visited on 12/02/2019).

Holt, L. L., A. J. Lotto, and K. R. Kluender (2001). "Influence of Fundamental Frequency on Stop-Consonant Voicing Perception: A Case of Learned Covariation or Auditory Enhancement?" In: The Journal of the Acoustical Society of America 109.2, pp. 764–774. issn: 0001-4966. doi: 10.1121/1.1339825. url: https://asa.scitation.org/doi/10.1121/1.1339825 (visited on 07/23/2021).

Hsieh, D.-L. et al. (2009). "Hearing in Noise Test in Subjects With Conductive Hearing Loss". In: Journal of the Formosan Medical Association 108.12, pp. 937–942. issn: 09296646. doi: 10.1016/S0929-6646(10)60006-X. url: https://linkinghub.elsevier.com/retrieve/pii/S092966461060006X (visited on 12/02/2019).

Hu, Y. et al. (2007). "Use of a Sigmoidal-Shaped Function for Noise Attenuation in Cochlear Implants". In: The Journal of the Acoustical Society of America 122.4, EL128–134. issn: 1520-8524. doi: 10.1121/1.2772401. pmid: 17902741.

Huang, J. et al. (2017). "Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise". In: Scientific Reports 7.1 (1), p. 2196. issn: 2045-2322. doi:

10.1038/s41598-017-02429-1. url: https://www.nature.com/articles/s41598-017-02429-1 (visited on 08/11/2021).

Huang, J. et al. (2019). "Electro-Tactile Stimulation Enhances Cochlear-Implant Melody Recognition: Effects of Rhythm and Musical Training". In: Ear & Hearing 41.1, pp. 106–113. issn: 0196-0202. doi: 10.1097/AUD.0000000000000749. url: https://journals.lww.com/00003446-202001000-00011 (visited on 08/11/2021).

Humes, L. E. and L. Roberts (1990). "Speech-Recognition Difficulties of the Hearing-Impaired Elderly: The Contributions of Audibility". In: p. 10.

Humes, L. E. and J. R. Dubno (2010). "Factors Affecting Speech Understanding in Older Adults". In: The Aging Auditory System. Ed. by S. Gordon-Salant et al. Vol. 34. New York, NY: Springer New York, pp. 211–257. isbn: 978-1-4419-0992-3 978-1-4419-0993-0. doi: 10.1007/978-1-4419-0993-0_8. url: http://link.springer.com/10.1007/978-1-4419-0993-0_8 (visited on 12/02/2019).

Humes, L. E., G. R. Kidd, and J. J. Lentz (2013). "Auditory and Cognitive Factors Underlying Individual Differences in Aided Speech-Understanding among Older Adults". In: Frontiers in Systems Neuroscience 7. issn: 1662-5137. doi: 10.3389/fnsys.2013.00055. url: http://journal.frontiersin.org/article/10.3389/fnsys.2013.00055/abstract (visited on 12/02/2019).

Hygge, S. et al. (1992). "Normal-Hearing and Hearing- Impaired Subjects' Ability to Just Follow Conversation in Competing Speech, Reversed Speech, and Noise Backgrounds". In: Journal of Speech, Language, and Hearing Research 35.1, pp. 208–215. doi: 10.1044/jshr.3501.208. url: https://pubs.asha.org/doi/abs/10.1044/jshr.3501.208 (visited on 11/28/2021).

Ifukube, T. and C. Yoshimoto (1974). "A Sono-Tactile Deaf-Aid Made of Piezoelectric Vibrator Array". In: The Journal of the Acoustical Society of Japan 30.8. url: https://www.jstage.jst.go.jp/article/jasj/30/8/30_KJ00001453098/_pdf/-char/ja (visited on 09/23/2021).

International Organization for Standardization (2001). ISO 13091-1:2001 - Mechanical Vibration - Vibrotactile Perception Thresholds for the Assessment of Nerve Dysfunction - Part 1: Methods of Measurement at the Fingertips. url: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/25/32509.html (visited on 10/04/2021).

International Telecomunication Union (2015). "Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level". In: ITU-R Recommendations. ITU-R Recommendations, p. 25. url: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf.

IRCAM (2018). Max Sound Box. Version 2018-3. Paris, FR: IRCAM. url: https://forum.ircam.fr/projects/detail/max-sound-box/ (visited on 10/01/2021).

Johnson, K. L. et al. (2008). "Developmental Plasticity in the Human Auditory Brainstem". In: Journal of Neuroscience 28.15, pp. 4000–4007. issn: 0270-6474, 1529-2401.

doi: 10.1523/JNEUROSCI.0012-08.2008. url: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0012-08.2008 (visited on 12/02/2019).

Johnson, K. O. and J. R. Phillips (1981). "Tactile Spatial Resolution. I. Two-point Discrimination, Gap Detection, Grating Resolution, and Letter Recognition". In: Journal of Neurophysiology 46.6, pp. 1177–1192. issn: 0022-3077, 1522-1598. doi: 10.1152/jn.1981.46.6.1177. url: https://www.physiology.org/doi/10.1152/jn.1981.46.6.1177 (visited on 09/21/2021).

Johnson, K. L., T. G. Nicol, and N. Kraus (2005). "Brain Stem Response to Speech: A Biological Marker of Auditory Processing:" in: Ear and Hearing 26.5, pp. 424–434. issn: 0196-0202. doi: 10.1097/01.aud.0000179687.71662.6e. url: https://insights.ovid.com/crossref?an=00003446-200510000-00002 (visited on 12/02/2019).

Jouvet, D. and Y. Laprie (2017). "Performance Analysis of Several Pitch Detection Algorithms on Simulated and Real Noisy Speech Data". In: 2017 25th European Signal Processing Conference (EUSIPCO). 2017 25th European Signal Processing Conference (EUSIPCO). Kos, Greece: IEEE, pp. 1614–1618. isbn: 978-0-9928626-7-1. doi: 10.23919/EUSIPCO.2017.8081482. url: http://ieeexplore.ieee.org/document/8081482/ (visited on 10/11/2021).

Juravle, G., F. McGlone, and C. Spence (2013). "Context-Dependent Changes in Tactile Perception during Movement Execution". In: Frontiers in Psychology 4, p. 913. issn: 1664-1078. doi: 10.3389/fpsyg.2013.00913. url: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00913 (visited on 09/29/2021).

Kaandorp, M. W. et al. (2015). "Assessing Speech Recognition Abilities with Digits in Noise in Cochlear Implant and Hearing Aid Users". In: International Journal of Audiology 54.1, pp. 48–57. issn: 1499-2027, 1708-8186. doi: 10.3109/14992027.2014.945623. url: http://www.tandfonline.com/doi/full/10.3109/14992027.2014.945623 (visited on 09/07/2021).

Kaernbach, C. and C. Bering (2001). "Exploring the Temporal Mechanism Involved in the Pitch of Unresolved Harmonics". In: The Journal of the Acoustical Society of America 110.2, pp. 1039–1048. issn: 0001-4966. doi: 10.1121/1.1381535. pmid: 11519572.

Kang, R. et al. (2009). "Development and Validation of the University of Washington Clinical Assessment of Music Perception Test". In: Ear and Hearing 30.4, pp. 411–418. issn: 1538-4667. doi: 10.1097/AUD.0b013e3181a61bc0. pmid: 19474735.

Kanold, P. O. and E. D. Young (2001). "Proprioceptive Information from the Pinna Provides Somatosensory Input to Cat Dorsal Cochlear Nucleus". In: The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 21.19, pp. 7848–7858. issn: 1529-2401. pmid: 11567076.

Kanold, P. O., K. A. Davis, and E. D. Young (2011). "Somatosensory Context Alters Auditory Responses in the Cochlear Nucleus". In: Journal of Neurophysiology 105.3, pp. 1063–1070. issn: 1522-1598. doi: 10.1152/jn.00807.2010. pmid: 21178001.

Kassuba, T. et al. (2013). “Multisensory Interactions between Auditory and Haptic Object Recognition”. In: Cerebral Cortex 23.5, pp. 1097–1107. issn: 1047-3211. doi: 10.1093/cercor/bhs076. url: https://doi.org/10.1093/cercor/bhs076 (visited on 12/07/2021).

Kates, J. M. and K. H. Arehart (2021). “The Hearing-Aid Speech Perception Index (HASPI) Version 2”. In: Speech Communication 131, pp. 35–46. issn: 0167-6393. doi: 10.1016/j.specom.2020.05.001. url: https://www.sciencedirect.com/science/article/pii/S0167639320300431 (visited on 04/17/2022).

Keidser, G. et al. (2002). “The National Acoustic Laboratories (NAL) CDs of Speech and Noise for Hearing Aid Evaluation: Normative Data and Potential Applications”. In: doi: 10.1375/AUDI.24.1.16.31112.

Kidd, G. et al. (2008). “Informational Masking”. In: Auditory Perception of Sound Sources. Ed. by W. A. Yost, A. N. Popper, and R. R. Fay. Springer Handbook of Auditory Research. Boston, MA: Springer US, pp. 143–189. isbn: 978-0-387-71305-2. doi: 10.1007/978-0-387-71305-2_6. url: https://doi.org/10.1007/978-0-387-71305-2_6 (visited on 07/24/2021).

Kidd, S. A. and J. B. Kelly (1996). “Contribution of the Dorsal Nucleus of the Lateral Lemniscus to Binaural Responses in the Inferior Colliculus of the Rat: Interaural Time Delays”. In: Journal of Neuroscience 16.22, pp. 7390–7397. issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.16-22-07390.1996. pmid: 8929445. url: https://www.jneurosci.org/content/16/22/7390 (visited on 12/05/2021).

Kiefer, J. et al. (2001). “Comparison of Speech Recognition with Different Speech Coding Strategies (SPEAK, CIS, and ACE) and Their Relationship to Telemetric Measures of Compound Action Potentials in the Nucleus CI 24M Cochlear Implant System”. In: Audiology 40.1, pp. 32–42. issn: 0020-6091. doi: 10.3109/00206090109073098. url: https://www.tandfonline.com/doi/abs/10.3109/00206090109073098 (visited on 09/07/2021).

Killion, M. C. and P. A. Niquette (2000). “What Can the Pure-Tone Audiogram Tell Us about a Patient’s SNR Loss?” In: 53.3, p. 6.

Killion, M. C. et al. (2004). “Development of a Quick Speech-in-Noise Test for Measuring Signal-to-Noise Ratio Loss in Normal-Hearing and Hearing-Impaired Listeners”. In: The Journal of the Acoustical Society of America 116.4, pp. 2395–2405. issn: 0001-4966. doi: 10.1121/1.1784440. url: http://asa.scitation.org/doi/10.1121/1.1784440 (visited on 12/02/2019).

Killion, M. C. et al. (2016). “Hearing Aid Transducers”. In: Hearing Aids. Ed. by G. R. Popelka et al. Springer Handbook of Auditory Research. Cham: Springer International Publishing, pp. 59–92. isbn: 978-3-319-33036-5. doi: 10.1007/978-3-319-33036-5_3. url: https://doi.org/10.1007/978-3-319-33036-5_3 (visited on 08/25/2021).

Kim, M.-B. et al. (2014). “Effect of a Bluetooth-implemented Hearing Aid on Speech Recognition Performance: Subjective and Objective Measurement”. In: The Annals of

Otology, Rhinology, and Laryngology 123.6, pp. 395–401. issn: 0003-4894. doi: 10.1177/0003489414526847. pmid: 24687593.

Klatt, D. H. and L. C. Klatt (1990). "Analysis, Synthesis, and Perception of Voice Quality Variations among Female and Male Talkers". In: The Journal of the Acoustical Society of America 87.2, pp. 820–857. issn: 0001-4966. doi: 10.1121/1.398894. url: http://asa.scitation.org/doi/10.1121/1.398894 (visited on 12/02/2019).

Kochkin, S. (2000). "MarkeTrak V: "Why My Hearing Aids Are in the Drawer": The Consumers' Perspective". In: 53.2, p. 5.

Kong, Y.-Y., A. Mullangi, and N. Ding (2014). "Differential Modulation of Auditory Responses to Attended and Unattended Speech in Different Listening Conditions". In: Hearing Research 316, pp. 73–81. issn: 03785955. doi: 10.1016/j.heares.2014.07.009. url: https://linkinghub.elsevier.com/retrieve/pii/S0378595514001270 (visited on 12/02/2019).

Kong, Y.-Y., A. Somarowthu, and N. Ding (2015). "Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech". In: Journal of the Association for Research in Otolaryngology 16.6, pp. 783–796. issn: 1525-3961, 1438-7573. doi: 10.1007/s10162-015-0540-x. url: http://link.springer.com/10.1007/s10162-015-0540-x (visited on 12/02/2019).

Kraus, N. and T. Nicol (2003). "Aggregate Neural Responses to Speech Sounds in the Central Auditory System". In: Speech Communication 41.1, pp. 35–47. issn: 01676393. doi: 10.1016/S0167-6393(02)00091-2. url: https://linkinghub.elsevier.com/retrieve/pii/S0167639302000912 (visited on 12/02/2019).

Kreft, H. A., G. S. Donaldson, and D. A. Nelson (2004). "Effects of Pulse Rate and Electrode Array Design on Intensity Discrimination in Cochlear Implant Users". In: The Journal of the Acoustical Society of America 116 (4 Pt 1), pp. 2258–2268. issn: 0001-4966. doi: 10.1121/1.1786871. pmid: 15532657.

Kreft, H. A., D. A. Nelson, and A. J. Oxenham (2013). "Modulation Frequency Discrimination with Modulated and Unmodulated Interference in Normal Hearing and in Cochlear-Implant Users". In: JARO: Journal of the Association for Research in Otolaryngology 14.4, pp. 591–601. issn: 1525-3961. doi: 10.1007/s10162-013-0391-2. pmid: 23632651. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705089/ (visited on 10/11/2021).

Krishnan, A. and J. T. Gandour (2009). "The Role of the Auditory Brainstem in Processing Linguistically-Relevant Pitch Patterns". In: Brain and Language 110.3, pp. 135–148. issn: 0093934X. doi: 10.1016/j.bandl.2009.03.005. url: https://linkinghub.elsevier.com/retrieve/pii/S0093934X0900042X (visited on 12/02/2019).

Lakatos, P. et al. (2007). "Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex". In: Neuron 53.2, pp. 279–292. issn: 0896-6273. doi: 10.1016/j.neuron.2006.12.011. pmid: 17224408.

Lalor, E. C. and J. J. Foxe (2010). "Neural Responses to Uninterrupted Natural Speech Can Be Extracted with Precise Temporal Resolution". In: European Journal of Neuroscience 31.1, pp. 189–193. issn: 0953816X, 14609568. doi: 10.1111/j.1460-9568.2009.07055.x. url: http://doi.wiley.com/10.1111/j.1460-9568.2009.07055.x (visited on 12/02/2019).

Lalor, E. C. et al. (2009). "Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli". In: Journal of Neurophysiology 102.1, pp. 349–359. issn: 0022-3077, 1522-1598. doi: 10.1152/jn.90896.2008. url: https://www.physiology.org/doi/10.1152/jn.90896.2008 (visited on 12/02/2019).

Larsby, B. and S. Arlinger (1994). "Speech Recognition and Just-Follow-Conversation Tasks for Normal-Hearing and Hearing-Impaired Listeners with Different Maskers". In: Audiology 33.3, pp. 165–176. issn: 0020-6091. doi: 10.3109/00206099409071877. url: https://www.tandfonline.com/doi/abs/10.3109/00206099409071877 (visited on 11/28/2021).

Laszig, R. et al. (2004). "Benefits of Bilateral Electrical Stimulation with the Nucleus Cochlear Implant in Adults: 6-Month Postoperative Results". In: Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 25.6, pp. 958–968. issn: 1531-7129. doi: 10.1097/00129492-200411000-00016. pmid: 15547426.

Launer, S., J. A. Zakis, and B. C. J. Moore (2016). "Hearing Aid Signal Processing". In: Hearing Aids. Ed. by G. R. Popelka et al. Springer Handbook of Auditory Research. Cham: Springer International Publishing, pp. 93–130. isbn: 978-3-319-33036-5. doi: 10.1007/978-3-319-33036-5_4. url: https://doi.org/10.1007/978-3-319-33036-5_4 (visited on 08/25/2021).

Laurienti, P. J. et al. (2006). "Enhanced Multisensory Integration in Older Adults". In: Neurobiology of Aging 27.8, pp. 1155–1163. issn: 01974580. doi: 10.1016/j.neurobiolaging.2005.05.024. url: https://linkinghub.elsevier.com/retrieve/pii/S0197458005001600 (visited on 08/15/2021).

Lawson, G. and M. Peterson (2011). Speech Audiometry. Core Clinical Concepts in Audiology. San Diego: Plural Pub. 166 pp. isbn: 978-1-59756-370-3.

Lee, L. W. and L. E. Humes (1993). "Evaluating a Speech-reception Threshold Model for Hearing-impaired Listeners". In: The Journal of the Acoustical Society of America 93.5, pp. 2879–2885. issn: 0001-4966. doi: 10.1121/1.405807. url: http://asa.scitation.org/doi/10.1121/1.405807 (visited on 12/02/2019).

Lee, S., A. Potamianos, and S. Narayanan (1999). "Acoustics of Children's Speech: Developmental Changes of Temporal and Spectral Parameters". In: The Journal of the Acoustical Society of America 105.3, pp. 1455–1468. issn: 0001-4966. doi: 10.1121/1.426686. pmid: 10089598.

Leek, M. R. (2001). "Adaptive Procedures in Psychophysical Research". In: Perception & Psychophysics 63.8, pp. 1279–1292. issn: 0031-5117, 1532-5962. doi: 10.3758/

BF03194543. url: http://link.springer.com/10.3758/BF03194543 (visited on 12/02/2019).

Leigh-Paffenroth, E. D. and O. D. Murnane (2011). "Auditory Steady State Responses Recorded in Multitalker Babble". In: International Journal of Audiology 50.2, pp. 86–97. issn: 1499-2027, 1708-8186. doi: 10.3109/14992027.2010.532512. url: http://www.tandfonline.com/doi/full/10.3109/14992027.2010.532512 (visited on 12/02/2019).

Lenarz, T. (2018). "Cochlear Implant – State of the Art". In: GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery 16, Doc04. issn: 1865-1011. doi: 10.3205/cto000143. pmid: 29503669. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5818683/ (visited on 09/06/2021).

Lesenfants, D. (2019). "Predicting Individual Speech Intelligibility from the Cortical Tracking of Acoustic- and Phonetic-Level Speech Representations". In: p. 20.

Lévêque, J. L. et al. (2000). "Changes in Tactile Spatial Discrimination and Cutaneous Coding Properties by Skin Hydration in the Elderly". In: The Journal of Investigative Dermatology 115.3, pp. 454–458. issn: 0022-202X. doi: 10.1046/j.1523-1747.2000.00055.x. pmid: 10951283.

Lewis, J. and B. Moss (2013). "MEMS Microphones, the Future for Hearing Aids". In: Analog Dialogue, p. 3.

Libenson, M. H. (2010). Practical Approach to Electroencephalography. Philadelphia, Pa: Saunders Elsevier. 335 pp. isbn: 978-0-7506-7478-2.

Loizou, P. C., A. Lobo, and Y. Hu (2005). "Subspace Algorithms for Noise Reduction in Cochlear Implants". In: The Journal of the Acoustical Society of America 118.5, pp. 2791–2793. issn: 0001-4966. pmid: 16334894. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1343472/ (visited on 09/06/2021).

Lopez-Poveda, E. A. et al. (2020). "Speech-in-Noise Recognition With More Realistic Implementations of a Binaural Cochlear-Implant Sound Coding Strategy Inspired by the Medial Olivocochlear Reflex". In: Ear and Hearing 41.6, pp. 1492–1510. issn: 1538-4667. doi: 10.1097/AUD.0000000000000880. pmid: 33136626.

Luo, X. and L. Hayes (2019). "Vibrotactile Stimulation Based on the Fundamental Frequency Can Improve Melodic Contour Identification of Normal-Hearing Listeners With a 4-Channel Cochlear Implant Simulation". In: Frontiers in Neuroscience 0. issn: 1662-453X. doi: 10.3389/fnins.2019.01145. url: https://www.frontiersin.org/articles/10.3389/fnins.2019.01145/full (visited on 08/11/2021).

Luts, H. et al. (2014). "Development and Normative Data for the Flemish/Dutch Matrix Test". In: International Journal of Audiology. url: https://www.semanticscholar.org/paper/Development-and-normative-data-for-the-Matrix-test-Luts-Jansen/4cdbfa3dc4e3139553b44b6a632665fd9456635a (visited on 09/03/2021).

Maddox, R. K. and A. K. C. Lee (2018). "Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners". In: eneuro 5.1, ENEURO.0441–17.2018. issn:

2373-2822. doi: 10.1523/ENEURO.0441-17.2018. url: http://eneuro.org/lookup/doi/10.1523/ENEURO.0441-17.2018 (visited on 12/02/2019).

Magnusson, L. (2011). "Comparison of the Fine Structure Processing (FSP) Strategy and the CIS Strategy Used in the MED-EL Cochlear Implant System: Speech Intelligibility and Music Sound Quality". In: International Journal of Audiology 50.4, pp. 279–287. issn: 1708-8186. doi: 10.3109/14992027.2010.537378. pmid: 21190508.

Majdak, P., C. Hollomey, and R. Baumgartner (2021). "AMT 1.0: The Toolbox for Reproducible Research in Auditory Modeling". In: ISAAR. url: https://amtoolbox.org/notes/MajdakHollomeyBaumgartner2021.pdf (visited on 04/17/2022).

Makhdoum, M. J. et al. (1998). "Intra- and Interindividual Correlations Between Auditory Evoked Potentials and Speech Perception in Cochlear Implant Users". In: Scandinavian Audiology 27.1, pp. 13–20. issn: 0105-0397. doi: 10.1080/010503998419650. url: http://www.tandfonline.com/doi/full/10.1080/010503998419650 (visited on 12/02/2019).

Manju, V., K. K. Gopika, and P. M. Arivudai Nambi (2014). "Association of Auditory Steady State Responses with Perception of Temporal Modulations and Speech in Noise". In: ISRN Otolaryngology 2014, pp. 1–8. issn: 2090-5750. doi: 10.1155/2014/374035. url: https://www.hindawi.com/archive/2014/374035/ (visited on 12/02/2019).

Mattys, S. L. et al. (2012). "Speech Recognition in Adverse Conditions: A Review". In: Language and Cognitive Processes 27.7-8, pp. 953–978. issn: 0169-0965. doi: 10.1080/01690965.2012.705006. url: https://doi.org/10.1080/01690965.2012.705006 (visited on 07/20/2021).

Mauger, S. J., K. Arora, and P. W. Dawson (2012). "Cochlear Implant Optimized Noise Reduction". In: Journal of Neural Engineering 9.6, p. 065007. issn: 1741-2552. doi: 10.1088/1741-2560/9/6/065007. pmid: 23187159.

McArdle, R. A., R. H. Wilson, and C. A. Burks (2005). "Speech Recognition in Multitalker Babble Using Digits, Words, and Sentences". In: Journal of the American Academy of Audiology 16.9, pp. 726–739. issn: 10500545. doi: 10.3766/jaaa.16.9.9. url: http://openurl.ingenta.com/content/xref?genre=article&issn=1050-0545&volume=16&issue=9&spage=726 (visited on 12/02/2019).

McCarthy, J. J. (1988). "Feature Geometry and Dependency: A Review". In: Phonetica 45.2-4, pp. 84–108. issn: 0031-8388, 1423-0321. doi: 10.1159/000261820. url: https://www.karger.com/Article/FullText/261820 (visited on 07/23/2021).

McDermott, H. J. (2004). "Music Perception with Cochlear Implants: A Review". In: Trends in Amplification 8.2, pp. 49–82. issn: 1084-7138. doi: 10.1177/108471380400800203. url: http://journals.sagepub.com/doi/10.1177/108471380400800203 (visited on 10/02/2021).

McDermott, J. H., M. Schemitsch, and E. P. Simoncelli (2013). "Summary Statistics in Auditory Perception". In: Nature Neuroscience 16.4 (4), pp. 493–498. issn: 1546-1726.

doi: 10.1038/nn.3347. url: https://www.nature.com/articles/nn.3347 (visited on 10/11/2021).

Mehta, A. H. and A. J. Oxenham (2017). "Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users". In: Journal of the Association for Research in Otolaryngology 18.6, pp. 789–802. issn: 1525-3961, 1438-7573. doi: 10.1007/s10162-017-0632-x. pmid: 28733803. url: http://link.springer.com/10.1007/s10162-017-0632-x (visited on 12/02/2019).

Meister, H. et al. (2009). "The Perception of Prosody and Speaker Gender in Normal-Hearing Listeners and Cochlear Implant Recipients". In: International Journal of Audiology 48.1, pp. 38–48. issn: 1708-8186. doi: 10.1080/14992020802293539. pmid: 19173112.

Merchel, S. and M. E. Altinsoy (2014). "The Influence of Vibrations on Musical Experience". In: Journal of the Audio Engineering Society 62.4, pp. 220–234. issn: 15494950. doi: 10.17743/jaes.2014.0016. url: http://www.aes.org/e-lib/browse.cfm?elib=17134 (visited on 11/30/2021).

Meredith, M. A. and B. L. Allman (2015). "Single-Unit Analysis of Somatosensory Processing in Core Auditory Cortex of Hearing Ferrets". In: The European journal of neuroscience 41.5, pp. 686–698. issn: 0953-816X. doi: 10.1111/ejn.12828. pmid: 25728185. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4347953/ (visited on 09/20/2021).

Middelweerd, M. J., J. M. Festen, and R. Plomp (1990). "Difficulties with Speech Intelligibility in Noise in Spite of a Normal Pure-Tone Audiogram: Original Papers". In: International Journal of Audiology 29.1, pp. 1–7. issn: 1499-2027, 1708-8186. doi: 10.3109/00206099009081640. url: http://www.tandfonline.com/doi/full/10.3109/00206099009081640 (visited on 12/02/2019).

Miller, J. and R. Ulrich (2001). "On the Analysis of Psychometric Functions: The Spearman-Kärber Method". In: Perception & Psychophysics 63.8, pp. 1399–1420. issn: 0031-5117, 1532-5962. doi: 10.3758/BF03194551. url: http://link.springer.com/10.3758/BF03194551 (visited on 12/02/2019).

Montoya-Martínez, J. et al. (2021). "Effect of Number and Placement of EEG Electrodes on Measurement of Neural Tracking of Speech". In: PLOS ONE 16.2, e0246769. issn: 1932-6203. doi: 10.1371/journal.pone.0246769. url: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0246769 (visited on 11/25/2021).

Moore, B. C. (1973). "Frequency Difference Limens for Short-Duration Tones". In: The Journal of the Acoustical Society of America 54.3, pp. 610–619. issn: 0001-4966. doi: 10.1121/1.1913640. pmid: 4754385.

Moore, B. C. and B. R. Glasberg (1988). "Gap Detection with Sinusoids and Noise in Normal, Impaired, and Electrically Stimulated Ears". In: The Journal of the Acoustical Society of America 83.3, pp. 1093–1101. issn: 0001-4966. doi: 10.1121/1.396054. pmid: 3356814.

Moore, B. C. J. (2016). An Introduction to the Psychology of Hearing. 6th ed. Emerald Group Publishing Limited, Bingley, UK. isbn: 97B-90-04-25M2-4.

Moore, B. C. J. and G. R. Popelka (2016). “Introduction to Hearing Aids”. In: Hearing Aids. Ed. by G. R. Popelka et al. Springer Handbook of Auditory Research. Cham: Springer International Publishing, pp. 1–19. isbn: 978-3-319-33036-5. doi: 10.1007/978-3-319-33036-5_1. url: https://doi.org/10.1007/978-3-319-33036-5_1 (visited on 08/25/2021).

Moradi, S. et al. (2014). “Gated Auditory Speech Perception: Effects of Listening Conditions and Cognitive Capacity”. In: Frontiers in Psychology 5. issn: 1664-1078. doi: 10.3389/fpsyg.2014.00531. url: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00531/abstract (visited on 12/02/2019).

Most, T. and M. Peled (2007). “Perception of Suprasegmental Features of Speech by Children With Cochlear Implants and Children With Hearing Aids”. In: The Journal of Deaf Studies and Deaf Education 12.3, pp. 350–361. issn: 1081-4159. doi: 10.1093/deafed/enm012. url: https://doi.org/10.1093/deafed/enm012 (visited on 09/28/2021).

MOTU (2019). MOTU 24Ao. Cambridge, MA, USA.

Müller, J. et al. (2012). “Clinical Trial Results with the MED-EL Fine Structure Processing Coding Strategy in Experienced Cochlear Implant Users”. In: ORL; journal for oto-rhino-laryngology and its related specialties 74.4, pp. 185–198. issn: 1423-0275. doi: 10.1159/000337089. pmid: 22814383.

Murray, I. R. and J. L. Arnott (1993). “Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion”. In: The Journal of the Acoustical Society of America 93.2, pp. 1097–1108. issn: 0001-4966. doi: 10.1121/1.405558. url: https://asa.scitation.org/doi/abs/10.1121/1.405558 (visited on 09/28/2021).

Nelson, P. B. et al. (2003). “Understanding Speech in Modulated Interference: Cochlear Implant Users and Normal-Hearing Listeners”. In: The Journal of the Acoustical Society of America 113.2, pp. 961–968. issn: 0001-4966. doi: 10.1121/1.1531983. url: http://asa.scitation.org/doi/10.1121/1.1531983 (visited on 09/06/2021).

Neosensory (2020). Buzz Tech Sheet. url: https://neosensory.com/wp-content/uploads/2020/10/Neosensory-Buzz-Tech-Sheet.pdf (visited on 08/12/2021).

Neupane, A. K. et al. (2014). “Effect of Repetition Rate on Speech Evoked Abr in Younger and Middle Aged Individuals”. In: Audiology Research 4.1. issn: 2039-4349, 2039-4330. doi: 10.4081/audiores.2014.106. url: http://www.audiologyresearch.org/index.php/audio/article/view/106 (visited on 12/02/2019).

Niedermeyer, E. and F. H. Lopes da Silva, eds. (2005). Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. 5th ed. Philadelphia: Lippincott Williams & Wilkins. 1309 pp. isbn: 978-0-7817-5126-1.

“NIH Consensus Conference. Cochlear Implants in Adults and Children” (1995). In: JAMA 274.24, pp. 1955–1961. issn: 0098-7484. pmid: 8568992.

Nilsson, M., S. D. Soli, and J. A. Sullivan (1994). "Development of the Hearing In Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise". In: The Journal of the Acoustical Society of America 95.2, pp. 1085–1099. issn: 0001-4966. doi: 10.1121/1.408469. url: http://asa.scitation.org/doi/10.1121/1.408469 (visited on 12/02/2019).

Niquette, P. A. and M. C. Killion (2016). BKB SIN User Manual.

Nordqvist, P. and A. Leijon (2004). "An Efficient Robust Sound Classification Algorithm for Hearing Aids". In: The Journal of the Acoustical Society of America 115.6, pp. 3033–3041. issn: 0001-4966. doi: 10.1121/1.1710877. url: https://asa.scitation.org/doi/10.1121/1.1710877 (visited on 09/05/2021).

Norris, D. and J. M. McQueen (2008). "Shortlist B: A Bayesian Model of Continuous Speech Recognition". In: Psychological Review 115.2, pp. 357–395. issn: 1939-1471(Electronic),0033-295X(Print). doi: 10.1037/0033-295X.115.2.357.

Novich, S. D. and D. M. Eagleman (2015). "Using Space and Time to Encode Vibrotactile Information: Toward an Estimate of the Skin's Achievable Throughput". In: Experimental Brain Research 233.10, pp. 2777–2788. issn: 1432-1106. doi: 10.1007/s00221-015-4346-1. pmid: 26080756.

Novis, K. and S. Bell (2019). "Objective Comparison of the Quality and Reliability of Auditory Brainstem Response Features Elicited by Click and Speech Sounds". In: Ear and Hearing 40.3, pp. 447–457. issn: 1538-4667. doi: 10.1097/AUD.0000000000000639. url: https://journals.lww.com/ear-hearing/Fulltext/2019/05000/Objective_Comparison_of_the_Quality_and.2.aspx (visited on 12/07/2021).

O'Sullivan, A. E., C. Y. Lim, and E. C. Lalor (2019). "Look at Me When I'm Talking to You: Selective Attention at a Multisensory Cocktail Party Can Be Decoded Using Stimulus Reconstruction and Alpha Power Modulations". In: The European Journal of Neuroscience 50.8, pp. 3282–3295. issn: 1460-9568. doi: 10.1111/ejn.14425. pmid: 31013361.

O'Sullivan, A. E. et al. (2021). "Neurophysiological Indices of Audiovisual Speech Processing Reveal a Hierarchy of Multisensory Integration Effects". In: The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 41.23, pp. 4991–5003. issn: 1529-2401. doi: 10.1523/JNEUROSCI.0906-20.2021. pmid: 33824190.

Okada, K. et al. (2010). "Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech". In: Cerebral Cortex 20.10, pp. 2486–2495. issn: 1460-2199, 1047-3211. doi: 10.1093/cercor/bhp318. url: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp318 (visited on 12/02/2019).

Oldfield, S. R. and S. P. A. Parker (1984). "Acuity of Sound Localisation: A Topography of Auditory Space. II. Pinna Cues Absent". In: Perception 13.5, pp. 601–617. issn: 0301-0066. doi: 10.1068/p130601. url: https://doi.org/10.1068/p130601 (visited on 04/16/2022).

Oshima, K. et al. (2010). “Curing Hearing Loss: Patient Expectations, Health Care Practitioners, and Basic Science”. In: Journal of communication disorders 43.4, pp. 311–318. issn: 0021-9924. doi: 10.1016/j.jcomdis.2010.04.002. pmid: 20434163. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885475/ (visited on 08/30/2021).

Oxenham, A. J. (2008). “Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants”. In: Trends in Amplification 12.4, pp. 316–331. issn: 1084-7138. doi: 10.1177/1084713808325881. pmid: 18974203. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901529/ (visited on 09/28/2021).

Oxenham, A. J. and S. P. Bacon (2003). “Cochlear Compression: Perceptual Measures and Implications for Normal and Impaired Hearing:” in: Ear and Hearing 24.5, pp. 352–366. issn: 0196-0202. doi: 10.1097/01.AUD.0000090470.73934.78. url: https://insights.ovid.com/crossref?an=00003446-200310000-00002 (visited on 12/02/2019).

Paludetti, G. et al. (1991). “Relationships between Middle Latency Auditory Responses (MLR) and Speech Discrimination Tests in the Elderly”. In: Acta Oto-Laryngologica 111 (sup476), pp. 105–109. issn: 0001-6489, 1651-2251. doi: 10.3109/00016489109127262. url: http://www.tandfonline.com/doi/full/10.3109/00016489109127262 (visited on 12/02/2019).

Parette, P. and M. Scherer (2004). “Assistive Technology Use and Stigma”. In: Education and Training in Developmental Disabilities 39.3, pp. 217–226. issn: 1547-0350. JSTOR: 23880164.

Parise, C. V. and M. O. Ernst (2016). “Correlation Detection as a General Mechanism for Multisensory Integration”. In: Nature Communications 7.1, p. 11543. issn: 2041-1723. doi: 10.1038/ncomms11543. url: http://www.nature.com/articles/ncomms11543 (visited on 08/15/2021).

Pedregosa, F. et al. (2011). “Scikit-Learn: Machine Learning in Python”. In: Journal of Machine Learning Research 12.85, pp. 2825–2830. issn: 1533-7928. url: http://jmlr.org/papers/v12/pedregosa11a.html (visited on 10/22/2021).

Pekkarinen, E., A. Salmivalli, and J. Suonpää (1990). “Effect of Noise on Word Discrimination by Subjects with Impaired Hearing, Compared with Those with Normal Hearing”. In: Scandinavian Audiology 19.1, pp. 31–36. issn: 0105-0397. doi: 10.3109/01050399009070749. url: http://www.tandfonline.com/doi/full/10.3109/01050399009070749 (visited on 12/02/2019).

Peng, S.-C., J. B. Tomblin, and C. W. Turner (2008). “Production and Perception of Speech Intonation in Pediatric Cochlear Implant Recipients and Individuals with Normal Hearing”. In: Ear & Hearing 29.3, pp. 336–351. issn: 0196-0202. doi: 10.1097/AUD.0b013e318168d94d. url: https://journals.lww.com/00003446-200806000-00004 (visited on 09/28/2021).

Penner, M. J. (1977). "Detection of Temporal Gaps in Noise as a Measure of the Decay of Auditory Sensation". In: The Journal of the Acoustical Society of America 61.2, pp. 552–557. issn: 0001-4966. doi: 10.1121/1.381297. url: https://asa.scitation.org/doi/10.1121/1.381297 (visited on 09/21/2021).

Penner, M. J. et al. (1974). "Intensity Discrimination for Pulsed Sinusoids of Various Frequencies". In: Perception & Psychophysics 15.3, pp. 568–570. issn: 1532-5962. doi: 10.3758/BF03199303. url: https://doi.org/10.3758/BF03199303 (visited on 09/21/2021).

Perez, C. A., C. A. Holzmann, and H. E. Jaeschke (2000). "Two-Point Vibrotactile Discrimination Related to Parameters of Pulse Burst Stimulus". In: Medical and Biological Engineering and Computing 38.1, pp. 74–79. issn: 1741-0444. doi: 10.1007/BF02344692. url: https://doi.org/10.1007/BF02344692 (visited on 09/21/2021).

Perez, C. A., J. R. Donoso, and L. E. Medina (2010). "A Critical Experimental Study of the Classical Tactile Threshold Theory". In: BMC Neuroscience 11.1, p. 76. issn: 1471-2202. doi: 10.1186/1471-2202-11-76. url: https://doi.org/10.1186/1471-2202-11-76 (visited on 09/29/2021).

Perry, S. W., S. L. Bell, and D. Simpson (2018). "Comparative Study of Methods for Predicting Speech-in-Noise Performance Using Evoked Responses". In: British Society of Audiology Basic Auditory Science. Newcastle. url: https://conferences.ncl.ac.uk/bas2018/AbstractsBookBAS2018.pdf.

– (2019). "Development of a Method for Predicting Speech-in-noise Performance Using Evoked Responses to Running Speech". In: British Society of Audiology Basic Auditory Science. University College London. url: https://www.ucl.ac.uk/ear/sites/ear/files/bas2019_programme.pdf.

Peters, R. W., B. C. Moore, and T. Baer (1998). "Speech Reception Thresholds in Noise with and without Spectral and Temporal Dips for Hearing-Impaired and Normally Hearing People". In: The Journal of the Acoustical Society of America 103.1, pp. 577–587. issn: 0001-4966. doi: 10.1121/1.421128. pmid: 9440343.

Pezent, E. et al. (2019). "Tasbi: Multisensory Squeeze and Vibrotactile Wrist Haptics for Augmented and Virtual Reality". In: 2019 IEEE World Haptics Conference (WHC). 2019 IEEE World Haptics Conference (WHC). Tokyo, Japan: IEEE, pp. 1–6. isbn: 978-1-5386-9461-9. doi: 10.1109/WHC.2019.8816098. url: https://ieeexplore.ieee.org/document/8816098/ (visited on 08/12/2021).

Picou, E. M., E. Aspell, and T. A. Ricketts (2014). "Potential Benefits and Limitations of Three Types of Directional Processing in Hearing Aids". In: Ear and Hearing 35.3, pp. 339–352. issn: 1538-4667. doi: 10.1097/AUD.0000000000000004. pmid: 24518429.

Picou, E. M., S. C. Marcrum, and T. A. Ricketts (2015). "Evaluation of the Effects of Nonlinear Frequency Compression on Speech Recognition and Sound Quality for Adults with Mild to Moderate Hearing Loss". In: International Journal of Audiology 54.3, pp. 162–169. issn: 1708-8186. doi: 10.3109/14992027.2014.961662. pmid: 25731581.

Picton, T. W. et al. (2003). “Human Auditory Steady-State Responses: Respuestas Auditivas de Estado Estable En Humanos”. In: International Journal of Audiology 42.4, pp. 177–219. issn: 1499-2027, 1708-8186. doi: 10.3109/14992020309101316. url: http://www.tandfonline.com/doi/full/10.3109/14992020309101316 (visited on 12/02/2019).

Plack, C. J. (2014). The Sense of Hearing. Second edition. New York: Psychology Press, Taylor & Francis Group. 296 pp. isbn: 978-1-84872-987-2 978-1-84872-515-7.

Plomp, R. (1964). “Rate of Decay of Auditory Sensation”. In: doi: 10.1121/1.1918946.

Plomp, R. and A. M. Mimpen (1979). “Improving the Reliability of Testing the Speech Reception Threshold for Sentences”. In: International Journal of Audiology 18.1, pp. 43–52. issn: 1499-2027, 1708-8186. doi: 10.3109/00206097909072618. url: http://www.tandfonline.com/doi/full/10.3109/00206097909072618 (visited on 12/02/2019).

Plomp, R. (1986). “A Signal-to-Noise Ratio Model for the Speech-Reception Threshold of the Hearing Impaired”. In: Journal of Speech, Language, and Hearing Research 29.2, pp. 146–154. issn: 1092-4388, 1558-9102. doi: 10.1044/jshr.2902.146. url: http://pubs.asha.org/doi/10.1044/jshr.2902.146 (visited on 12/02/2019).

Power, A. J., R. B. Reilly, and E. C. Lalor (2011). “Comparing Linear and Quadratic Models of the Human Auditory System Using EEG”. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Boston, MA: IEEE, pp. 4171–4174. isbn: 978-1-4577-1589-1 978-1-4244-4121-1 978-1-4244-4122-8. doi: 10.1109/IEMBS.2011.6091035. url: http://ieeexplore.ieee.org/document/6091035/ (visited on 09/03/2021).

Precision Microdrives (2021a). 304-116 Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/304-116-5mm-vibration-motor-20mm-type-datasheet.pdf (visited on 09/23/2021).

– (2021b). 306-10H Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/306-10h-7mm-vibration-motor-25mm-type-datasheet.pdf (visited on 09/23/2021).

– (2021c). 307-103 Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/307-103-9mm-vibration-motor-25mm-type-datasheet.pdf (visited on 09/23/2021).

– (2021d). C08-005 Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/c08-005-8mm-linear-resonant-actuator-3mm-type-datasheet.pdf (visited on 09/23/2021).

– (2021e). C08-00A Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/c08-00a-8mm-linear-resonant-actuator-3mm-type-datasheet.pdf (visited on 09/23/2021).

Precision Microdrives (2021f). C10-100 Datasheet. url: https://www.precisionmicrodrives.com/product/datasheet/c10-100-10mm-linear-resonant-actuator-4mm-type-datasheet.pdf (visited on 09/23/2021).

Proctor, A. and M. H. Goldstein (1983). "Development of Lexical Comprehension in a Profoundly Deaf Child Using a Wearable, Vibrotactile Communication Aid". In: Language, Speech, and Hearing Services in Schools 14.3, pp. 138–149. doi: 10.1044/0161-1461.1403.138. url: https://pubs.asha.org/doi/abs/10.1044/0161-1461.1403.138 (visited on 09/21/2021).

Rawool, V. W. (1989). "Speech Recognition Scores and ABR in Cochlear Impairment". In: Scandinavian Audiology 18.2, pp. 113–117. issn: 0105-0397. doi: 10.3109/01050398909070731. url: http://www.tandfonline.com/doi/full/10.3109/01050398909070731 (visited on 12/02/2019).

Reichenbach, C. S. et al. (2016). "The Auditory-Brainstem Response to Continuous, Non-repetitive Speech Is Modulated by the Speech Envelope and Reflects Speech Processing". In: Frontiers in Computational Neuroscience 10. issn: 1662-5188. doi: 10.3389/fncom.2016.00047. url: http://journal.frontiersin.org/Article/10.3389/fncom.2016.00047/abstract (visited on 12/02/2019).

Ricketts, T. A. and B. W. Y. Hornsby (2005). "Sound Quality Measures for Speech in Noise through a Commercial Hearing Aid Implementing Digital Noise Reduction". In: Journal of the American Academy of Audiology 16.5, pp. 270–277. issn: 1050-0545. doi: 10.3766/jaaa.16.5.2. pmid: 16119254.

Riecke, L. et al. (2019). "Audio-Tactile Enhancement of Cortical Speech-Envelope Tracking". In: NeuroImage 202. issn: 10538119. doi: 10.1016/j.neuroimage.2019.116134. url: https://linkinghub.elsevier.com/retrieve/pii/S1053811919307256 (visited on 12/11/2020).

Riss, D. et al. (2011). "Envelope Versus Fine Structure Speech Coding Strategy: A Crossover Study". In: Otology & Neurotology 32.7, pp. 1094–1101. issn: 1531-7129. doi: 10.1097/MAO.0b013e31822a97f4. url: https://journals.lww.com/otology-neurotology/Abstract/2011/09000/Envelope_Versus_Fine_Structure_Speech_Coding.10.aspx (visited on 09/07/2021).

Riss, D. et al. (2016). "Effects of Stimulation Rate With the FS4 and HDCIS Coding Strategies in Cochlear Implant Recipients". In: Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 37.7, pp. 882–888. issn: 1537-4505. doi: 10.1097/MAO.0000000000001107. pmid: 27295444.

Rivolta, M. N. (2013). "New Strategies for the Restoration of Hearing Loss: Challenges and Opportunities". In: British Medical Bulletin 105.1, pp. 69–84. issn: 0007-1420. doi: 10.1093/bmb/lds035. url: https://doi.org/10.1093/bmb/lds035 (visited on 08/30/2021).

RME (n.d.). RME Babyface Pro. Haimhausen, Germany: RME. url: https://babyface.rme-audio.de/#.

Romero, A. et al. (2015). “Auditory Middle Latency Response and Phonological Awareness in Students with Learning Disabilities”. In: International Archives of Otorhinolaryngology 19.04, pp. 325–330. issn: 1809-9777, 1809-4864. doi: 10.1055/s-0035-1551552. url: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0035-1551552 (visited on 12/02/2019).

Rosen, S. et al. (1992). “Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects”. In: Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 336.1278, pp. 367–373. doi: 10.1098/rstb.1992.0070. url: https://royalsocietypublishing.org/doi/10.1098/rstb.1992.0070 (visited on 07/23/2021).

Rosen, S. et al. (2013). “Listening to Speech in a Background of Other Talkers: Effects of Talker Number and Noise Vocoding”. In: The Journal of the Acoustical Society of America 133.4, pp. 2431–2443. issn: 0001-4966. doi: 10.1121/1.4794379. url: http://asa.scitation.org/doi/10.1121/1.4794379 (visited on 12/02/2019).

Rothenberg, M. et al. (1977). “Vibrotactile Frequency for Encoding a Speech Parameter”. In: The Journal of the Acoustical Society of America 62.4, pp. 1003–1012. issn: 0001-4966. doi: 10.1121/1.381610. pmid: 908786.

Russo, F. A., P. Ammirante, and D. I. Fels (2012). “Vibrotactile Discrimination of Musical Timbre.” In: Journal of Experimental Psychology: Human Perception and Performance 38.4, pp. 822–826. issn: 1939-1277, 0096-1523. doi: 10.1037/a0029046. url: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0029046 (visited on 11/30/2021).

Russo, N. et al. (2004). “Brainstem Responses to Speech Syllables”. In: Clinical Neurophysiology 115.9, pp. 2021–2030. issn: 13882457. doi: 10.1016/j.clinph.2004.04.003. url: https://linkinghub.elsevier.com/retrieve/pii/S1388245704001440 (visited on 12/02/2019).

Schatzle, S. et al. (n.d.). “Evaluation of Vibrotactile Feedback to the Human Arm”. In: (), p. 4.

Schnupp, J., I. Nelken, and A. King (2011). Auditory Neuroscience: Making Sense of Sound. Cambridge, Mass: MIT Press. 356 pp. isbn: 978-0-262-11318-2.

Schoof, T. and S. Rosen (2014). “The Role of Auditory and Cognitive Factors in Understanding Speech in Noise by Normal-Hearing Older Listeners”. In: Frontiers in Aging Neuroscience 6. issn: 1663-4365. doi: 10.3389/fnagi.2014.00307. url: http://journal.frontiersin.org/article/10.3389/fnagi.2014.00307/abstract (visited on 12/02/2019).

Schürmann, M. et al. (2006). “Touch Activates Human Auditory Cortex”. In: NeuroImage 30.4, pp. 1325–1331. issn: 1053-8119. doi: 10.1016/j.neuroimage.2005.11.020. pmid: 16488157.

Scott, S. K. (2000). “Identification of a Pathway for Intelligible Speech in the Left Temporal Lobe”. In: Brain 123.12, pp. 2400–2406. issn: 14602156. doi: 10.1093/brain/123.12.2400. url: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/123.12.2400 (visited on 12/02/2019).

Sek, A. and B. C. J. Moore (1995). “Frequency Discrimination as a Function of Frequency, Measured in Several Ways”. In: The Journal of the Acoustical Society of America 97.4, pp. 2479–2486. issn: 0001-4966. doi: 10.1121/1.411968. url: https://asa.scitation.org/doi/10.1121/1.411968 (visited on 12/07/2021).

Shackleton, T. M. and R. P. Carlyon (1994). “The Role of Resolved and Unresolved Harmonics in Pitch Perception and Frequency Modulation Discrimination”. In: The Journal of the Acoustical Society of America 95.6, pp. 3529–3540. issn: 0001-4966. doi: 10.1121/1.409970. pmid: 8046144.

Shannon, R. V., F. G. Zeng, and J. Wygonski (1998). “Speech Recognition with Altered Spectral Distribution of Envelope Cues”. In: The Journal of the Acoustical Society of America 104.4, pp. 2467–2476. issn: 0001-4966. doi: 10.1121/1.423774. pmid: 10491708.

Shannon, R. V. et al. (1995). “Speech Recognition with Primarily Temporal Cues”. In: Science 270.5234. Cited By :2039, pp. 303–304. issn: 0036-8075, 1095-9203. doi: 10.1126/science.270.5234.303. pmid: 7569981. url: http://www.sciencemag.org/cgi/doi/10.1126/science.270.5234.303 (visited on 12/02/2019).

Shannon, R. V. (1990). “Forward Masking in Patients with Cochlear Implants”. In: The Journal of the Acoustical Society of America 88.2, pp. 741–744. issn: 0001-4966. doi: 10.1121/1.399777. url: https://asa.scitation.org/doi/abs/10.1121/1.399777 (visited on 09/21/2021).

Shibasaki, M., Y. Kamiyama, and K. Minamizawa (2016). “Designing a Haptic Feedback System for Hearing-Impaired to Experience Tap Dance”. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology. UIST ’16 Adjunct. New York, NY, USA: Association for Computing Machinery, pp. 97–99. isbn: 978-1-4503-4531-6. doi: 10.1145/2984751.2985716. url: https://doi.org/10.1145/2984751.2985716 (visited on 11/29/2021).

Shield, B. M. and J. E. Dockrell (2008). “The Effects of Environmental and Classroom Noise on the Academic Attainments of Primary School Children”. In: The Journal of the Acoustical Society of America 123.1, pp. 133–144. issn: 0001-4966. doi: 10.1121/1.2812596. url: http://asa.scitation.org/doi/10.1121/1.2812596 (visited on 12/02/2019).

Shore, S. E., H. El Kashlan, and J. Lu (2003). “Effects of Trigeminal Ganglion Stimulation on Unit Activity of Ventral Cochlear Nucleus Neurons”. In: Neuroscience 119.4, pp. 1085–1101. issn: 0306-4522. doi: 10.1016/s0306-4522(03)00207-0. pmid: 12831866.

Shore, S. E. et al. (2000). “Trigeminal Ganglion Innervates the Auditory Brainstem”. In: The Journal of Comparative Neurology 419.3, pp. 271–285. issn: 0021-9967. doi: 10.1002/(sici)1096-9861(20000410)419:3<271::aid-cne1>3.0.co;2-m. pmid: 10723004.

Simpson, A. (2009). "Frequency-Lowering Devices for Managing High-Frequency Hearing Loss: A Review". In: Trends in Amplification 13.2, pp. 87–106. issn: 1084-7138. doi: 10.1177/1084713809336421. pmid: 19447764.

Simpson, S. A. and M. Cooke (2005). "Consonant Identification in N-talker Babble Is a Nonmonotonic Function of N". In: The Journal of the Acoustical Society of America 118.5, pp. 2775–2778. issn: 0001-4966. doi: 10.1121/1.2062650. url: http://asa.scitation.org/doi/10.1121/1.2062650 (visited on 12/02/2019).

Skinner, M. W. et al. (2002). "Speech Recognition with the Nucleus 24 SPEAK, ACE, and CIS Speech Coding Strategies in Newly Implanted Adults". In: Ear and Hearing 23.3, pp. 207–223. issn: 0196-0202. doi: 10.1097/00003446-200206000-00005. pmid: 12072613.

Skoe, E. and N. Kraus (2010). "Auditory Brain Stem Response to Complex Sounds: A Tutorial:" in: Ear and Hearing 31.3, pp. 302–324. issn: 0196-0202. doi: 10.1097/AUD.0b013e3181cdb272. url: https://insights.ovid.com/crossref?an=00003446-201006000-00002 (visited on 12/02/2019).

Smiljanić, R. and A. R. Bradlow (2009). "Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes". In: Language and linguistics compass 3.1, pp. 236–264. issn: 1749-818X. doi: 10.1111/j.1749-818X.2008.00112.x. pmid: 20046964. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2747755/ (visited on 09/03/2021).

Smoorenburg, G. F. (1992). "Speech Reception in Quiet and in Noisy Conditions by Individuals with Noise-induced Hearing Loss in Relation to Their Tone Audiogram". In: The Journal of the Acoustical Society of America 91.1, pp. 421–437. issn: 0001-4966. doi: 10.1121/1.402729. url: http://asa.scitation.org/doi/10.1121/1.402729 (visited on 12/02/2019).

Smulders, Y. E. et al. (2016). "Comparison of Bilateral and Unilateral Cochlear Implantation in Adults: A Randomized Clinical Trial". In: JAMA Otolaryngology–Head & Neck Surgery 142.3, pp. 249–256. issn: 2168-6181. doi: 10.1001/jamaoto.2015.3305. url: https://doi.org/10.1001/jamaoto.2015.3305 (visited on 09/06/2021).

Soli, S. D. (2008). "Some Thoughts on Communication Handicap and Hearing Impairment". In: International Journal of Audiology 47.6, pp. 285–286. issn: 1499-2027, 1708-8186. doi: 10.1080/14992020802060656. url: http://www.tandfonline.com/doi/full/10.1080/14992020802060656 (visited on 12/02/2019).

Somers, B., C. J. Long, and T. Francart (2021). "EEG-based Diagnostics of the Auditory System Using Cochlear Implant Electrodes as Sensors". In: Scientific Reports 11.1 (1), p. 5383. issn: 2045-2322. doi: 10.1038/s41598-021-84829-y. url: https://www.nature.com/articles/s41598-021-84829-y (visited on 11/25/2021).

Somers, B., E. Verschueren, and T. Francart (2019). "Neural Tracking of the Speech Envelope in Cochlear Implant Users". In: Journal of Neural Engineering 16.1, p. 016003. issn: 1741-2560, 1741-2552. doi: 10.1088/1741-2552/aae6b9. url: https://iopscience.iop.org/article/10.1088/1741-2552/aae6b9 (visited on 09/30/2020).

Song, J. et al. (2006). “On the Relationship between Speech- and Nonspeech-Evoked Auditory Brainstem Responses”. In: Audiology and Neurotology 11.4, pp. 233–241. issn: 1420-3030, 1421-9700. doi: 10.1159/000093058. url: https://www.karger.com/Article/FullText/93058 (visited on 12/02/2019).

Souza, P. (2016). “Speech Perception and Hearing Aids”. In: Hearing Aids. Ed. by G. R. Popelka et al. Springer Handbook of Auditory Research. Cham: Springer International Publishing, pp. 151–180. isbn: 978-3-319-33036-5. doi: 10.1007/978-3-319-33036-5_6. url: https://doi.org/10.1007/978-3-319-33036-5_6 (visited on 08/25/2021).

Space Systems. Definition of the Technology Readiness Levels (TRLs) and Their Criteria of Assessment (2013). Space Systems. Definition of the Technology Readiness Levels (TRLs) and Their Criteria of Assessment: BSI British Standards. doi: 10.3403/30260201. url: https://linkresolver.bsigroup.com/junction/resolve/000000000030260201?restype=standard (visited on 04/10/2022).

Spens, K.-E. and G. Plant (1984). “A Tactual ’Hearing’ Aid for the Deaf”. In: Proceedings of the Tenth International Congress of Phonetic Sciences. De Gruyter Mouton, pp. 733–737. isbn: 978-3-11-088468-5. doi: 10.1515/9783110884685-121. url: https://www.degruyter.com/document/doi/10.1515/9783110884685-121/pdf (visited on 09/21/2021).

Sperry, J. L., T. L. Wiley, and M. R. Chial (n.d.). “Word Recognition Performance in Various Background Competitors”. In: (), p. 10.

Spieth, W., J. F. Curtis, and J. C. Webster (1954). “Responding to One of Two Simultaneous Messages”. In: The Journal of the Acoustical Society of America 26.3, pp. 391–396. issn: 0001-4966. doi: 10.1121/1.1907347. url: http://asa.scitation.org/doi/10.1121/1.1907347 (visited on 12/02/2019).

Spriet, A. et al. (2007). “Speech Understanding in Background Noise with the Two-Microphone Adaptive Beamformer BEAM in the Nucleus Freedom Cochlear Implant System”. In: Ear and Hearing 28.1, pp. 62–72. issn: 0196-0202. doi: 10.1097/01.aud.0000252470.54246.54. pmid: 17204899.

Stevens, K. N. (2002). “Toward a Model for Lexical Access Based on Acoustic Landmarks and Distinctive Features”. In: The Journal of the Acoustical Society of America 111.4, pp. 1872–1891. issn: 0001-4966. doi: 10.1121/1.1458026. url: https://asa.scitation.org/doi/10.1121/1.1458026 (visited on 07/23/2021).

Stone, M. A. et al. (2008). “Tolerable Hearing Aid Delays. V. Estimation of Limits for Open Canal Fittings”. In: Ear and Hearing 29.4, pp. 601–617. issn: 1538-4667. doi: 10.1097/AUD.0b013e3181734ef2. pmid: 18469715.

Sullivan, J. R. and C. Carrano (2015). “Working Memory and Speech Recognition Performance in Noise: Implications for Classroom Accommodations”. In: Journal of Communication Disorders, Deaf Studies & Hearing Aids 03.03. issn: 23754427. doi: 10.4172/2375-4427.1000136. url: http://www.esciencecentral.org/journals/working-memory-and-speech-recognition-performance-in-noise-implicationsforclassroom-accommodations-2375-4427-1000136.php?aid=55947 (visited on 12/02/2019).

Summers, I. et al. (2005). "Tactile Information Transfer: A Comparison of Two Stimulation Sites". In: The Journal of the Acoustical Society of America 118, pp. 2527–34. doi: 10.1121/1.2031979.

Summers, I. R. (1992). Tactile Aids for the Hearing Impaired. Wiley. 272 pp. isbn: 978-1-870332-17-0. Google Books: Wu17QgAACAAJ.

Summers, V. and M. R. Molis (2004). "Speech Recognition in Fluctuating and Continuous Maskers: Effects of Hearing Loss and Presentation Level". In: p. 13.

Suzuki, Y. and H. Takeshima (2004). "Equal-Loudness-Level Contours for Pure Tones". In: The Journal of the Acoustical Society of America 116.2, pp. 918–933. issn: 0001-4966. doi: 10.1121/1.1763601. pmid: 15376658.

Synigal, S. R., E. S. Teoh, and E. C. Lalor (2020). "Including Measures of High Gamma Power Can Improve the Decoding of Natural Speech From EEG". In: Frontiers in Human Neuroscience 14, p. 130. issn: 1662-5161. doi: 10.3389/fnhum.2020.00130. pmid: 32410969.

Tactaid II+ Fitting Manual M-3 (1992).

Taylor, B. (2003). "Speech-in-Noise Tests: How and Why to Include Them in Your Basic Test Battery". In: The Hearing Journal 56.1, p. 40. issn: 0745-7472. doi: 10.1097/01.HJ.0000293000.76300.ff. url: https://journals.lww.com/thehearingjournal/Fulltext/2003/01000/Speech_in_noise_tests__How_and_why_to_include_them.8.aspx (visited on 12/05/2021).

TDK (2021a). PowerHap 0904H014V060 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/0904h014v060.pdf.

– (2021b). PowerHap 0909H011V060 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/0909h011v060.pdf.

– (2021c). PowerHap 1204H018V060 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/1204h018v060.pdf.

– (2021d). PowerHap 1313H018V120 Datasheet.

– (2021e). PowerHap 1919H021V120 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/1919h021v120.pdf.

– (2021f). PowerHap 2626H023V120 Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/2626h023v120.pdf.

– (2021g). PowerHap 6005H070V120 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/6005h070v120.pdf.

TDK (2021h). PowerHap 6005H090V120 Prototype Datasheet. url: https://product.tdk.com/system/files/dam/doc/product/sw_piezo/haptic/powerhap/data_sheet/20/10/ds/6005h090v120.pdf.

Teoh, E. S., M. S. Cappelloni, and E. C. Lalor (2019). "Prosodic Pitch Processing Is Represented in Delta-Band EEG and Is Dissociable from the Cortical Tracking of Other Acoustic and Phonetic Features". In: European Journal of Neuroscience 50.11, pp. 3831–3842. issn: 1460-9568. doi: 10.1111/ejn.14510. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.14510 (visited on 10/30/2021).

Teoh, E. S. and E. C. Lalor (2019). "EEG Decoding of the Target Speaker in a Cocktail Party Scenario: Considerations Regarding Dynamic Switching of Talker Location". In: Journal of Neural Engineering 16.3, p. 036017. issn: 1741-2552. doi: 10.1088/1741-2552/ab0cf1. pmid: 30836345.

Texas Instruments (2021). DRV2605L 2- to 5.2-V Haptic Driver for LRA and ERM with Effect Library and Smart-Loop Architecture Datasheet. url: https://www.ti.com/lit/ds/symlink/drv2605l.pdf (visited on 09/23/2021).

The Mathwork Inc. (2019). MATLAB. Version 2019a. Natick, MA, USA.

Theunissen, M., D. W. Swanepoel, and J. Hanekom (2009). "Sentence Recognition in Noise: Variables in Compilation and Interpretation of Tests". In: International Journal of Audiology 48.11, pp. 743–757. issn: 1499-2027, 1708-8186. doi: 10.3109/14992020903082088. url: http://www.tandfonline.com/doi/full/10.3109/14992020903082088 (visited on 12/02/2019).

Titze, I. R. (1989). "Physiologic and Acoustic Differences between Male and Female Voices". In: The Journal of the Acoustical Society of America 85.4, pp. 1699–1707. issn: 0001-4966. doi: 10.1121/1.397959. pmid: 2708686.

Tranchant, P. et al. (2017). "Feeling the Beat: Bouncing Synchronization to Vibrotactile Music in Hearing and Early Deaf People". In: Frontiers in Neuroscience 11, p. 507. issn: 1662-4548. doi: 10.3389/fnins.2017.00507. pmid: 28955193. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5601036/ (visited on 11/30/2021).

Turgeon, C. et al. (2015). "Deficits in Auditory Frequency Discrimination and Speech Recognition in Cochlear Implant Users". In: Cochlear Implants International 16.2, pp. 88–94. issn: 1754-7628. doi: 10.1179/1754762814Y.0000000091. pmid: 25117940.

Uchanski, R. M. (2005). "Clear Speech". In: The Handbook of Speech Perception. John Wiley & Sons, Ltd, pp. 207–235. isbn: 978-0-470-75702-4. doi: 10.1002/9780470757024.ch9. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470757024.ch9 (visited on 09/03/2021).

Vallat, R. (2018). "Pingouin: Statistics in Python". In: Journal of Open Source Software 3.31, p. 1026. issn: 2475-9066. doi: 10.21105/joss.01026. url: https://joss.theoj.org/papers/10.21105/joss.01026 (visited on 10/22/2021).

Van Engen, K. J. and A. R. Bradlow (2007). "Sentence Recognition in Native- and Foreign-Language Multi-Talker Background Noise". In: The Journal of the Acoustical

Society of America 121.1, pp. 519–526. issn: 0001-4966. doi: 10.1121/1.2400666. url: http://asa.scitation.org/doi/10.1121/1.2400666 (visited on 12/02/2019).

Van Eyndhoven, S., T. Francart, and A. Bertrand (2017). "EEG-Informed Attended Speaker Extraction From Recorded Speech Mixtures With Application in Neuro-Steered Hearing Prostheses". In: IEEE Transactions on Biomedical Engineering 64.5, pp. 1045–1056. issn: 1558-2531. doi: 10.1109/TBME.2016.2587382.

Vandali, A. et al. (2019). "Evaluation of the Optimized Pitch and Language Strategy in Cochlear Implant Recipients". In: Ear and Hearing 40.3, pp. 555–567. issn: 1538-4667. doi: 10.1097/AUD.0000000000000627. pmid: 30067558.

Vanheusden, F. J. et al. (2020). "Hearing Aids Do Not Alter Cortical Entrainment to Speech at Audible Levels in Mild-to-Moderately Hearing-Impaired Subjects". In: Frontiers in Human Neuroscience 14, p. 109. issn: 1662-5161. doi: 10.3389/fnhum.2020.00109. url: https://www.frontiersin.org/article/10.3389/fnhum.2020.00109/full (visited on 10/06/2020).

Vanthornhout, J., L. Decruy, and T. Francart (2019). "Effect of Task and Attention on Neural Tracking of Speech". In: Frontiers in Neuroscience 13, p. 977. issn: 1662-453X. doi: 10.3389/fnins.2019.00977. url: https://www.frontiersin.org/article/10.3389/fnins.2019.00977/full (visited on 02/08/2021).

Vanthornhout, J. et al. (2018). "Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope". In: Journal of the Association for Research in Otolaryngology 19.2, pp. 181–191. issn: 1525-3961, 1438-7573. doi: 10.1007/s10162-018-0654-z. url: http://link.springer.com/10.1007/s10162-018-0654-z (visited on 12/02/2019).

Verhey, J. L., J. Rennies, and S. M. A. Ernst (2007). "Influence of Envelope Distributions on Signal Detection". In: Acta Acustica united with Acustica 93.1, pp. 115–121.

Verrillo, R. T. (1963). "Effect of Contractor Area on the Vibrotactile Threshold". In: Journal of the Acoustical Society of America 35.12, pp. 1962–1971. issn: 0001-4966. doi: 10.1121/1.1918868.

Verrillo, R. T., A. J. Fraioli, and R. L. Smith (1969). "Sensation Magnitude of Vibrotactile Stimuli". In: Perception & Psychophysics 6.6, pp. 366–372. issn: 0031-5117, 1532-5962. doi: 10.3758/BF03212793. url: http://link.springer.com/10.3758/BF03212793 (visited on 09/21/2021).

Verschueren, E., B. Somers, and T. Francart (2019). "Neural Envelope Tracking as a Measure of Speech Understanding in Cochlear Implant Users". In: Hearing Research 373, pp. 23–31. issn: 0378-5955. doi: 10.1016/j.heares.2018.12.004. url: https://www.sciencedirect.com/science/article/pii/S0378595518304878 (visited on 10/30/2021).

Verschueren, E., J. Vanthornhout, and T. Francart (2020). "The Effect of Stimulus Choice on an EEG-Based Objective Measure of Speech Intelligibility". In: Ear and Hearing 41.6, pp. 1586–1597. issn: 0196-0202. doi: 10.1097/AUD.0000000000000875. url: https://journals.lww.com/ear-hearing/Abstract/2020/11000/The_Effect_of_Stimulus_Choice_on_an_EEG_Based.16.aspx (visited on 10/15/2021).

Virtanen, P. et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: Nature Methods 17.3 (3), pp. 261–272. issn: 1548-7105. doi: 10.1038/s41592-019-0686-2. url: https://www.nature.com/articles/s41592-019-0686-2 (visited on 10/22/2021).

Von Kriegstein, K., R. D. Patterson, and T. Griffiths (2008). “Task-Dependent Modulation of Medial Geniculate Body Is Behaviorally Relevant for Speech Recognition”. In: Current Biology 18.23, pp. 1855–1859. issn: 09609822. doi: 10.1016/j.cub.2008.10.052. url: https://linkinghub.elsevier.com/retrieve/pii/S0960982208014218 (visited on 12/02/2019).

Wagener, K. C. and T. Brand (2005). “Sentence Intelligibility in Noise for Listeners with Normal Hearing and Hearing Impairment: Influence of Measurement Procedure and Masking Parameters La Inteligibilidad de Frases En Silencio Para Sujetos Con Audición Normal y Con Hipoacusia: La Influencia Del Procedimiento de Medición y de Los Parámetros de Enmascaramiento”. In: International Journal of Audiology 44.3, pp. 144–156. issn: 1499-2027, 1708-8186. doi: 10.1080/14992020500057517. url: http://www.tandfonline.com/doi/full/10.1080/14992020500057517 (visited on 12/02/2019).

Wagner, M. et al. (2016). “Representation of Spectro-Temporal Features of Spoken Words within the P1-N1-P2 and T-complex of the Auditory Evoked Potentials (AEP)”. In: Neuroscience Letters 614, pp. 119–126. issn: 03043940. doi: 10.1016/j.neulet.2015.12.020. url: https://linkinghub.elsevier.com/retrieve/pii/S0304394015303128 (visited on 12/02/2019).

Walker, J., G. Fenn, and B. Smith (1987). “Soft-Talker: A Sound Level Monitor for the Hard-of-Hearing Using an Improved Tactile Transducer”. In: Journal of Biomedical Engineering 9.2, pp. 177–179. issn: 01415425. doi: 10.1016/0141-5425(87)90032-X. url: https://linkinghub.elsevier.com/retrieve/pii/014154258790032X (visited on 09/22/2021).

Wallace, M. T., L. K. Wilkinson, and B. E. Stein (1996). “Representation and Integration of Multiple Sensory Inputs in Primate Superior Colliculus”. In: Journal of Neurophysiology 76.2, pp. 1246–1266. issn: 0022-3077, 1522-1598. doi: 10.1152/jn.1996.76.2.1246. url: https://www.physiology.org/doi/10.1152/jn.1996.76.2.1246 (visited on 08/15/2021).

Walter, W. G. (1964). “The Convergence and Interaction of Visual, Auditory, and Tactile Responses in Human Nonspecific Cortex”. In: Annals of the New York Academy of Sciences 112.1, pp. 320–361. issn: 1749-6632. doi: 10.1111/j.1749-6632.1964.tb26760.x. url: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1964.tb26760.x (visited on 08/24/2021).

Wang, D. and J. H. L. Hansen (2018). “Speech Enhancement for Cochlear Implant Recipients”. In: The Journal of the Acoustical Society of America 143.4, p. 2244. issn: 1520-8524. doi: 10.1121/1.5031112. pmid: 29716262.

Weisenberger, J. M. (1986). “Sensitivity to Amplitude-Modulated Vibrotactile Signals”. In: The Journal of the Acoustical Society of America 80.6, pp. 1707–1715. issn: 0001-4966. doi: 10.1121/1.394283. pmid: 3794077.

Weisenberger, J. M. (1989). “Evaluation of the Siemens Minifonator Vibrotactile Aid”. In: Journal of Speech, Language, and Hearing Research 32.1, pp. 24–32. doi: 10.1044/jshr.3201.24. url: https://pubs.asha.org/doi/10.1044/jshr.3201.24 (visited on 08/12/2021).

Wessinger, C. M. et al. (2001). “Hierarchical Organization of the Human Auditory Cortex Revealed by Functional Magnetic Resonance Imaging”. In: Journal of Cognitive Neuroscience 13.1, pp. 1–7. issn: 0898-929X, 1530-8898. doi: 10.1162/089892901564108. url: http://www.mitpressjournals.org/doi/10.1162/089892901564108 (visited on 12/02/2019).

Whalen, D. H. et al. (1993). “F0 Gives Voicing Information Even with Unambiguous Voice Onset Times”. In: The Journal of the Acoustical Society of America 93.4, pp. 2152–2159. issn: 0001-4966. doi: 10.1121/1.406678. url: https://asa.scitation.org/doi/10.1121/1.406678 (visited on 07/23/2021).

Wilson, B. S. (2008). “Cochlear Implants: Current Designs and Future Possibilities”. In: The Journal of Rehabilitation Research and Development 45.5, pp. 695–730. issn: 07487711. doi: 10.1682/JRRD.2007.10.0173. url: http://www.rehab.research.va.gov/jour/08/45/5/pdf/wilson.pdf (visited on 04/16/2022).

– (2015). “Getting a Decent (but Sparse) Signal to the Brain for Users of Cochlear Implants”. In: Hearing Research. Lasker Award 322, pp. 24–38. issn: 0378-5955. doi: 10.1016/j.heares.2014.11.009. url: https://www.sciencedirect.com/science/article/pii/S0378595514001993 (visited on 10/11/2021).

Wilson, R. H., H. B. Abrams, and A. L. Pillion (2003). “A Word-Recognition Task in Multitalker Babble Using a Descending Presentation Mode from 24 dB to 0 dB Signal to Babble”. In: The Journal of Rehabilitation Research and Development 40.4, p. 321. issn: 0748-7711. doi: 10.1682/JRRD.2003.07.0321. url: http://www.rehab.research.va.gov/jour/03/40/4/pdf/Wilson-B.pdf (visited on 12/02/2019).

Wilson, R. H. and C. A. Burks (2005). “Use of 35 Words for Evaluation of Hearing Loss in Signal-to-Babble Ratio: A Clinic Protocol”. In: The Journal of Rehabilitation Research and Development 42.6, p. 839. issn: 0748-7711. doi: 10.1682/JRRD.2005.01.0009. url: http://www.rehab.research.va.gov/jour/05/42/6/pdf/wilson.pdf (visited on 12/02/2019).

Wilson, R. H., C. S. Carnell, and A. L. Cleghorn (2007). “The Words-in-Noise (WIN) Test with Multitalker Babble and Speech-Spectrum Noise Maskers”. In: Journal of the American Academy of Audiology 18.6, pp. 522–529. issn: 1050-0545. doi: 10.3766/jaaa.18.6.7. url: http://www.ingentaconnect.com/content/10.3766/jaaa.18.6.7 (visited on 12/02/2019).

Wilson, S. M., A. Bautista, and A. McCarron (2018). “Convergence of Spoken and Written Language Processing in the Superior Temporal Sulcus”. In: NeuroImage 171, pp. 62–74. issn: 1053-8119. doi: 10.1016/j.neuroimage.2017.12.068. url: https://www.sciencedirect.com/science/article/pii/S1053811917310923 (visited on 07/23/2021).

Wingfield, A. (1996). "Cognitive Factors in Auditory Performance: Context, Speed of Processing, and Constraints of Memory". In: Journal of the American Academy of Audiology 7.3, p. 8.

Wouters, J., H. McDermott, and T. Francart (2015). "Sound Coding in Cochlear Implants: From Electric Pulses to Hearing". In: IEEE Signal Processing Magazine 32, pp. 67–80. doi: 10.1109/MSP.2014.2371671.

Wright, S. J. (2015). "Coordinate Descent Algorithms". In: Mathematical Programming 151.1, pp. 3–34. issn: 1436-4646. doi: 10.1007/s10107-015-0892-3. url: https://doi.org/10.1007/s10107-015-0892-3 (visited on 10/22/2021).

Xin Luo, n., Q.-J. Fu, and J. J. Galvin (2007). "Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users". In: Trends in Amplification 11.4, pp. 301–315. issn: 1084-7138. doi: 10.1177/1084713807305301. pmid: 18003871.

Xu, L. and B. E. Pfingst (2008). "Spectral and Temporal Cues for Speech Recognition: Implications for Auditory Prostheses". In: Hearing research 242.1-2, pp. 132–140. issn: 0378-5955. doi: 10.1016/j.heares.2007.12.010. pmid: 18249077. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2610393/ (visited on 07/23/2021).

Yang, M. et al. (2015). "Speech Reconstruction from Human Auditory Cortex with Deep Neural Networks". In: Interspeech 2015. Interspeech 2015. ISCA, pp. 1121–1125. doi: 10.21437/Interspeech.2015-294. url: https://www.isca-speech.org/archive/interspeech_2015/yang15c_interspeech.html (visited on 10/30/2021).

Ye, H. et al. (2013). "A Wavelet-Based Noise Reduction Algorithm and Its Clinical Evaluation in Cochlear Implants". In: PLOS ONE 8.9, e75662. issn: 1932-6203. doi: 10.1371/journal.pone.0075662. url: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0075662 (visited on 09/06/2021).

Yueh, B. et al. (2003). "Screening and Management of Adult Hearing Loss in Primary Care: Scientific Review". In: JAMA 289.15, p. 1976. issn: 0098-7484. doi: 10.1001/jama.289.15.1976. url: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.289.15.1976 (visited on 07/22/2021).

Zeng, F.-G. and J. J. I. Galvin (1999). "Amplitude Mapping and Phoneme Recognition in Cochlear Implant Listeners". In: Ear and Hearing 20.1, pp. 60–74. issn: 0196-0202. url: https://journals.lww.com/ear-hearing/Abstract/1999/02000/Amplitude_Mapping_and_Phoneme_Recognition_in.6.aspx (visited on 09/12/2021).

Zeng, F.-G. et al. (2002). "Speech Dynamic Range and Its Effect on Cochlear Implant Performance". In: The Journal of the Acoustical Society of America 111.1, pp. 377–386. issn: 0001-4966. doi: 10.1121/1.1423926. url: http://asa.scitation.org/doi/10.1121/1.1423926 (visited on 09/12/2021).

Zhou, Q. et al. (2014). "A Reduction of the Elastic Net to Support Vector Machines with an Application to GPU Computing". In: p. 7.

Zou, H. and T. Hastie (2005). "Regularization and Variable Selection via the Elastic Net". In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67.2, pp. 301–320. issn: 1369-7412, 1467-9868. doi: 10.1111/j.1467-9868.2005.00503.x. url: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2005.00503.x (visited on 12/09/2021).

Zuk, N. J. et al. (2021). "Envelope Reconstruction of Speech and Music Highlights Stronger Tracking of Speech at Low Frequencies". In: PLoS computational biology 17.9, e1009358. issn: 1553-7358. doi: 10.1371/journal.pcbi.1009358. pmid: 34534211.

Statement on the impact of COVID-19 on this thesis

The COVID-19 pandemic has had a significant impact on this thesis. The restrictions imposed as a result of the COVID-19 pandemic has substantially impeded: collection of data for both a human neuroimaging study and behavioural study, due to the inherent close contact needed; data analysis, due to the adaptions required for remote analysis using the high-performance computing cluster; and physical testing and analysis of devices, due to the lack of access to critical equipment from the university's vibration centre (ISVR). The pandemic has had a clear impact on 3 studies across the thesis, with data collection shortened for two studies and one study cancelled entirely. Of the two shortened studies, it was only possible to collect pilot data for the first study and a reduced number of participants for the second. The result of this was a lack of novel data for assessment of the research questions, with a substantial adaptation of analysis protocol required for these studies. A number of adaptions have been implemented to mitigate the effects of COVID. This includes the design of 2 new studies using sub-optimal substitute datasets from other projects at the university. In addition to the full implementation of data analysis pipelines for the two new datasets, code for all studies required extensive adaption to run remotely on high-performance computing facilities (IRIDIS), given the reduced access to in-lab facilities. Lack of access to in-lab resources has also severely limited the physical characterization of haptic devices for the electro-haptics study. This has been mitigated somewhat by focusing instead on further optimisation to signal-processing strategies, however unimpeded access to the labs would have benefited the final thesis considerably. These mitigations, although effective for maintaining quality of thesis content, have required considerable time for implementation of revised data collection, data analysis and signal-processing strategy optimisation. The combination of adaptions listed have significantly altered the project's scope, and limited conclusions possible for the posed research questions. It is hoped that these exceptional circumstances and the efforts taken to mitigate the impact of COVID-19 will be taken into consideration.