# DECODING AUDITORY ATTENTION AND NEURAL LANGUAGE PROCESSING IN ADVERSE CONDITIONS AND DIFFERENT LISTENER GROUPS

*Anna Exenberger*

A thesis submitted in fulfilment of the requirements for the degree of

*Doctor of Philosophy*

**Department of Speech, Hearing and Phonetic Sciences**

University College London

2019

## Declaration

I, Anna Exenberger, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Anna Exenberger

## Acknowledgements

First and foremost, I would like to thank my supervisor Paul Iverson, for providing invaluable guidance, advice and help in a very hands-on way throughout this PhD. I have learned an awful lot about speech science, EEG testing and analysis and writing from him. I’ve also learned to use hyphens less, and I do sincerely hope that reading this thesis will be easier as a result of it. Thank you also to my secondary supervisor Stuart Rosen, and to Valerie Hazan, for your advice and guidance on my research and on the PhD process more generally. Special thanks to Outi Tuomainen, for your honesty, understanding and real help. Thank you also to Tobias Reichenbach and Octave Etard for their expertise and code to help with FFR analyses.

I would also like to thank Andrew Clark for his help setting up experiments and putting out fires big and small before and during testing. Thank you to Julia Habicht, Peter Derleth, Stefan Klockgether, Marlene Zippel, Nick Herbert and all the others from Sonova AG for their generous financial support and all the know-how they shared with me, making it possible to set up and conduct an entire study in Switzerland in a mere three months. Special thanks to Jieun Song, for your help understanding entrainment, permutation stats and lots of other things. A big thank you also to my colleagues in the SHaPS PhD room – Max Paulus, Gwijde Maegherman, Giulia Borghini, Julie Saigusa, Shiran Koifmann and everyone else that went for coffees with me and listened patiently to me complaining about my testing, MATLAB and life issues. I am very grateful to the ENRICH network and all my fellow PhDs – Katherine Marcoux, Amy Hall, Sneha Raman, and everyone else… we will all be butterflies one day!

Sincere thanks to the European Union for providing the financial support to establish such an international network through the Marie Curie Actions. Thank you to all the participants that I have tested over the past three years, for patiently waiting while I was setting up the EEG system, and for listening to countless sentences and stories in difficult noise.

I would like to thank my friends for staying my friends and always having a patient ear – Magdalena Schwarz, Christine Strobel, Marie Lang, Anna Miklas, Lisa Gruber, Flora Endel and everyone that spent some time with me over a cup of tea or a glass of wine. Thank you to Mat, who has made the last and final stretch of this PhD more enjoyable, and provided me with lots of iron-rich meals to combat fatigue. Thank you also to my family, for making my holidays at home a time to relax and rewind, and to take my mind off participants, papers and the psychology of hearing. I would not have made it through those years without your unconditional support and your belief that I can handle everything that's being thrown my way.

Thank you, dear reader, for making the time to read this thesis.

# Abstract

This thesis investigated subjective, behavioural and neurophysiological (EEG) measures of speech processing in various adverse conditions and with different listener groups. In particular, this thesis focused on different neural processing stages and their relationship with auditory attention, effort, and measures of speech intelligibility. Study 1 set the groundwork by establishing a toolbox of various neural measures to investigate online speech processing, from the frequency following response (FFR) and cortical measures of speech processing, to the N400, a measure of lexico-semantic processing. Results showed that peripheral processing is heavily influenced by stimulus characteristics such as degradation, whereas central processing units are more closely linked to higher-order phenomena such as speech intelligibility. In Study 2, a similar experimental paradigm was used to investigate differences in neural processing between a hearing-impaired and a normal-hearing group. Subjects were presented with short stories in different levels of multi-talker babble noise, and with different settings on their hearing aids. Findings indicate that, particularly at lower noise levels, the hearing-impaired group showed much higher cortical entrainment than the normal-hearing group, despite similar levels of speech recognition. Intersubject correlation, another global neural measure of auditory attention, however, was similarly affected by noise levels in both the hearing-impaired and the normal-hearing group. This finding indicates extra processing in the hearing-impaired group only on the level of the auditory cortex. Study 3, in contrast to Studies 1 and 2 (which both investigated the effects of bottom-up factors on neural processing), examined the links between entrainment and top-down factors, specifically motivation; as well as reasons for the

higher entrainment found in hearing-impaired subjects in Study 2. Results indicated that, while behaviourally there was no difference between incentive and non-incentive conditions, neurophysiological measures of attention such as intersubject correlation were affected by the presence of an incentive to perform better. Moreover, using a specific degradation type resulted in subjects' increased cortical entrainment under degraded conditions. These findings support the hypothesis that top-down factors such as motivation influence neurophysiological measures; and that higher entrainment to degraded speech might be triggered specifically by the reduced availability of spectral detail contained in speech.

# Impact statement

The work presented in this thesis benefits both specific industries such as hearing aid manufacturers and speech technology companies, as well as the general public and society.

## Industry partners

The insights gathered in this thesis on how hearing impairment affects listening effort and auditory attention are directly beneficial for the hearing aid industry. The large differences in neural entrainment between a normal-hearing and a hearing-impaired group found in the second study of this thesis suggest that neural language processing in older, hearing-impaired people is altered in significant ways compared to younger, normal-hearing people. These findings emphasise the importance of looking beyond simple sound amplification to mitigate hearing impairment. This particular study conducted in collaboration with Sonova AG introduced, for the first time, the use of electrophysiological measures to a wide audience ranging from engineers to audiologists within the company. The testing of different processing algorithms and their impact on neural measures can be expanded in the future with varied parameters (such as different SNRs or distractor types). The machine-learning tools used to decode auditory attention might be implemented in devices such as cognitively steered hearing aids, which provide selective amplification of the desired sound stream only.

Through the European ENRICH network, there has been close industry interaction with various network partners from speech technology and hearing aid companies as well as hospital clinics throughout this PhD. These involved biannual multi-day network meetings and project progress presentations, as well as special events such as entrepreneurship training and the presentation of innovative business products to industry partners (Crete 2017), or an industry networking event (Aachen 2019).

**General public and society**

According to the WHO, currently around 466 million people worldwide have disabling hearing loss, and this figure is predicted to double by 2050. Hearing aid users regularly report dissatisfaction with their devices, and the drop-out rate is high. Thus, the research conducted in this thesis benefits the general public and addresses the need for research and development in a still drastically underserved and underdeveloped market. The combination of different measures of auditory attention used in this thesis contributes to a growing body of research on listening effort, and supports the developments in clinical practice and industry towards quantifying the amount of effort expended for a given auditory task. Such tools, in turn, can contribute to the advancement of devices such as more effective hearing aids that are better suited to individual needs.

Additionally, this work is made accessible to the interested general public. Within the ENRICH network, there will be a public engagement event held at the Royal Institution in London (March 2020) with EEG demonstrations and the display and explanation of results from this thesis. Parts of this thesis have also been presented at the British Federation of Women Graduates (June 2019) to an audience consisting of female professors from different disciplines as well as fellow PhDs.

## Contents

**Chapter 3: Neurophysiological measures of attention in hearing aid users............................................................................................ 78**

# List of figures

# List of tables

# Chapter 1: General introduction

## 1.1 Speech processing in adverse conditions

Listening to speech in adverse conditions, such as in background noise, can be challenging. For example, when talking to someone in a crowded room with other talkers present, we need to be able to focus on one speaker while ignoring the others. This challenge of selective attention in a multi-talker environment is called the cocktail party effect (Cherry, 1953). While people with normal hearing are surprisingly good at speech recognition in cocktail party situations, the neural mechanisms underlying selective attention in difficult conditions are still poorly understood. This is not just a theoretical problem, but also a practical one: some listeners have been shown to have considerably more trouble attending to a speaker in cocktail party environments. Two example groups with challenges of extracting speech from a noisy background are hearing-impaired people and computer speech recognisers. Particularly for the hearing-impaired, the reduced ability to understand speech as well as the additional effort expended in difficult situations can have wide-ranging impacts on their personal lives, potentially leading to social withdrawal, feelings of isolation and loneliness (Bronkhorst & Plomp, 1992). For this reason, a better understanding of the mechanisms underlying speech perception and the factors governing it would allow for the development of more effective tools to mitigate against adverse effects for the ever-increasing global group of hearing-impaired people, as well as for machines recognizing speech.

Difficult listening conditions can be grouped into different classes, depending on the source and form of degradation. In a review paper, Mattys, Davis, Bradlow, & Scott (2012) outlined three main origins of adverse (difficult) listening conditions: (1) source degradation, e.g. accented speech, speech disorder, (2) environmental/transmission degradation, e.g. traffic noise, other talkers (also referred to as 'maskers'), and (3) receiver limitations, e.g. incomplete language model (non-native listener), or a hearing impairment. Different types of adverse conditions affect listening in different ways, and thus require different mitigations. For example, a noise reduction algorithm might be very helpful in a situation where someone is trying to have a conversation over the phone with a lot of background noise. The same algorithm, however, might be significantly less effective in a situation where someone is having a hard time understanding someone with an unfamiliar foreign accent, because the source of degradation is different. As an additional complication, in everyday situations there are often multiple types of adverse conditions present at the same time. Consequently, a better understanding of the distinct ways in which different types of adverse conditions affect listening can assist in the development of mitigation techniques that benefit both speaker and listener. This thesis investigated adverse conditions of all three types outlined above: chapter 2 presents a study that discusses source and environmental degradation, chapter 3 discusses hearing impairment as a form of receiver limitation, and chapter 4 presents a study with a motivation manipulation as a form of source manipulation.

### 1.1.1 Different types of degradation

Previous research has determined that one important factor determining the effectiveness of different speech maskers is their degree of similarity to the target signal (Rosen, Souza, Ekelund, & Majeed, 2013). Speech itself is a particularly good masker for other speech not just because of its acoustic similarity to the target signal, but also because it has the potential to draw attention away from the target on a semantic level. This distinction between purely peripheral effects of noise interference at the cochlear level and more central effects of distraction and attention is commonly referred to as the differentiating factor between energetic and informational masking (Brungart, 2001). While energetic masking has been well studied and theorized (see, for example, French & Steinberg, 1947 for the Articulation Index or Cooke, 2006 for the glimpsing model), the mechanisms of informational masking and competition for the listener's attention are less well researched (Zhang et al., 2016). Both types of masking can interfere with speech perception at the stages of auditory object formation as well as object selection (Shinn-Cunningham, 2008). While the failure to form separate auditory objects is due to the local structure being insufficient to separate one source from another (Kidd, Mason, & Arbogast, 2002), object selection can fail because the listener does not direct attention to the right object, for example because target and masker features are too similar to one another (Darwin, Brungart, & Simpson, 2003).

Speech in noise perception is most commonly measured on the power ratio between the signal and the noise (SNR), with positive SNRs indicating higher power in the target and negative SNRs indicating higher power in the noise signal. However, it is not always simple to determine the SNR levels corresponding to a set level of intelligibility. This is at least partly due to both target and masker signals usually being non-stationary, thus constantly changing the actual amount of target and masker power present in the current signal mix. Previous research has shown that a more accurate predictor of speech in noise recognition than global SNR levels is the degree of 'glimpsing' possible of the target signal. Glimpses occur in spectrotemporal regions where the target is least affected by noise (Cooke, 2006).

Performance on a speech-in-noise task is not just dependent on the type and level of noise (Rhebergen, Versfeld, & Dreschler, 2008; Wong, Ng, & Soli, 2012), but also on other determinants such as spatial cues and listener-specific factors. For example, a 90 degree spatial separation of target and masker improves speech recognition performance by up to 5 to 10dB and thus allows for the formation of separate auditory objects (Arbogast, Mason, & Kidd, 2005; Nilsson, Gelnett, Sullivan, Soli, & Goldberg, 1992; Peissig & Kollmeier, 1997). Furthermore, when listening in a non-native language, people are more affected by various types of noises compared to listening in their native language (Broersma & Scharenborg, 2010). There is also an abundance of research demonstrating that ageing and hearing impairment adversely affects people's speech perception skills in noise (Dubno, Dirks, & Morgan, 1984; Goossens, Vercammen, Wouters, & van Wieringen, 2017).

While a noise masker predominantly provides energetic masking and interferes at the stage of auditory object formation, an intelligible speech masker also adds informational masking by competing for the listener's attention, and can thus also impair object selection (Shinn-Cunningham, 2008). As a result of this competition, informational masking is assumed to operate on more central processes than energetic masking, which predominantly affects the auditory periphery (Brungart, 2001). A good example of failure at the auditory object selection stage is the experimental demonstration that listeners can often recall their names being presented to the non-target ear unexpectedly, even though they had been instructed to only attend to auditory input into the other (target) ear (Moray, 1959; Wood & Cowan, 1995). Further research with native listeners of different languages has found that people were less affected by maskers in other languages than their own (Lecumberri & Cooke, 2006; Rhebergen, Versfeld, & Dreschler, 2005). This indicates that even when speech is not attended, its linguistic content is processed to a certain degree, and that it can trigger involuntary attentional switches away from the target which result in impairment of auditory object selection.

As for energetic masking, informational masking can be reduced by spatial separation of the target and distractor, assisting with the formation of separate auditory objects. The degree to which spatial cues can provide a release from masking crucially depend on the similarity of target and distractor speakers. In an experiment with a female target speaker and two different maskers, namely speech-shaped noise and a male distractor talker, it was found that the release of masking from spatial separation was about the same for the noise as it was for the talker (Duquesnoy, 1983). However, importantly, in this experiment target and distractor speakers were of the opposite sex. When the same conditions were tested with a same-sex distractor speaker, it was found that the spatial separation advantage for the female talker amounted to 14 dB, which is almost double the 8dB release from masking measured when target and distractor speakers are of different sexes, or when the distractor is speech-shaped noise (Freyman, Helfer, McCall, & Clifton, 1999). These findings indicate that the benefit of spatial separation in informational masking conditions may only be accessed when needed, i.e. when the task is not easy enough to allow for the formation of separate auditory objects based on other cues such as fundamental frequency.

Apart from the possibility of adding noise and/or speech to a signal, the target itself can also be degraded. Vocoding as a commonly used technique, particularly in cochlear implants (CIs), divides the signal into a chosen number of frequency bands, while preserving most of the slowly varying temporal cues (temporal envelope) (Loizou, 2006). Vocoding has been used because it significantly reduces speech intelligibility compared to unprocessed speech in a systematic way, particularly in noisy conditions (Faulkner, Rosen, & Wilkinson, 2001; Friesen, Shannon, Baskent, & Wang, 2001; Qin & Oxenham, 2003). One of the main factors contributing to the reduced intelligibility of vocoded speech is the extremely weak, if not non-existent, representation of voice pitch (Oxenham, 2018). In CIs, the rectification stage of signal processing effectively removes temporal fine structure cues (Loizou, 2006). As temporal fine structure cues are known to be significant for pitch perception of complex harmonic sounds (Smith, Delgutte, & Oxenham, 2002), fundamental frequency (or its perceptual correlate voice pitch) is conveyed poorly in CIs.

Since fundamental frequency has been identified as an important cue for the segregation of speech streams in the presence of distractor talkers or background noise

(Başkent & Gaudrain, 2016; Brungart, 2001; Darwin et al., 2003), speech perception of vocoded stimuli is more difficult in noise, compared to stimuli where the f0 is preserved. Particularly in the presence of competing talkers, vocoding has been found to leave the listener more susceptible to informational masking, as pitch cues could not be used to discriminate the speakers (Stickney, Zeng, Litovsky, & Assmann, 2004).

### 1.1.2 Brain bases of speech processing

The effect of adverse conditions on speech processing can be investigated using different measures, such as speech recognition scores. However, these behavioural measures cannot provide direct insights into the brain bases and neural mechanisms underlying different performance levels in difficult listening situations. One non-invasive way of measuring online speech processing in the brain is the electroencephalography (EEG), which records electrical activity on the scalp generated by brain neurons. Using varying locations of placement of electrodes and different experimental paradigms, EEG as a comparatively inexpensive tool can be used to investigate speech processing in the brain from early brainstem responses to later cortical processing.

#### 1.1.2.1 The frequency following response

One of the earliest stages of speech processing in the brain is measurable in the brainstem, with a response latency of approximately 10ms. The auditory brainstem response (ABR) is an evoked potential recorded on the scalp, which is generated by the firing of many neurons in the brainstem in response to auditory stimulation. It consists of a sequence of peaks with different latencies, reflecting processing from the auditory nerve up to the inferior colliculus (Nuttall, Heinrich, Moore, & De Boer, 2013). The ABR can be recorded passively, i.e. without the subject's active doing. Click-evoked ABRs have thus been widely used in clinical settings to determine auditory thresholds and to detect neuropathologies (Hall, 2006). Since the auditory brainstem response is small compared to background noise (such as eye blinks, electrical artefacts and muscle/movement artefacts), in experiments it has typically been measured through averaging the response over many trials (Madsen, Harte, Elberling, & Dau, 2018).

The frequency following response (FFR), the sustained portion of the brainstem response, can be evoked by periodic stimuli and is believed to reflect the neural phase-locking to the fundamental frequency (f0) of human speech (Parbery-Clark, Strait, & Kraus, 2011; Skoe & Kraus, 2010; Song, Nicol, & Kraus, 2011). Its fidelity to the signal is high: in an experiment where the FFR to single words was recorded and the averaged response played back to subjects, they could identify the words with above-chance levels (Galbraith, Arbagey, Branski, Comerci, & Rector, 1995). The neural synchronization ensures that critical timing information contained in the speech signal is preserved.

Studies have demonstrated that processing in the brainstem is associated with successful speech perception, and that the FFR can be shaped by long-term experience and training. For example, studies with children have shown that poor readers and children with language-based learning problems show impaired brainstem encoding of speech sounds (Chandrasekaran, Hornickel, Skoe, Nicol, & Kraus, 2009; Russo, Nicol, Zecker, Hayes, & Kraus, 2005). In contrast, high school students with musical training were found to have enhanced neural language processing and better phonological awareness (Tierney, Krizman, & Kraus, 2015). Furthermore, speakers of tone languages (where pitch is used as a distinguishing semantic feature) and musicians with pitch training have more robust f0 representations, which is associated with better speech-in-noise performance (Bidelman, Gandour, & Krishnan, 2011; Parbery-Clark et al., 2011).

#### 1.1.2.2 Speech processing in auditory cortex and neural entrainment

Traditionally, as for brainstem level research, studies investigating speech processing at the cortical level have measured time-locked responses to the presentation of stimuli, averaged over a number of trials. This method provides an effective way of eliminating noise from other sources such as muscle artefacts and eye movements that impact the response on an individual trial level. These averaged, time-locked responses are referred to as event-related potentials (ERPs) and thought to reflect different stages in the sensory and cognitive processing of complex stimuli (Martin, Tremblay, & Korczak, 2008).

There are several well-known ERP responses to auditory stimuli generated at the cortical level. While earlier responses of up to roughly 100ms have been classified as ‘sensory’ (or ‘automatic) as they are primarily driven by physical stimulus features, later, ‘cognitive’ components are thought to be more reflective of the subject’s cognitive information processing (Sur & Sinha, 2009). Important parameters for the evaluation of any ERPs are the response latency and amplitude. For example, the P1-

N1-P2 complex (named after the approximate response time in milliseconds after stimulus presentation, i.e. 100-200ms) is understood to be an obligatory marker of cortical processing in response to auditory stimuli (Davis & Zerlin, 1966; Näätänen & Picton, 1987). Another well-researched phenomenon is the P300, often elicited through an oddball paradigm (Donchin & Coles, 1988; Friedman, Cycowicz, & Gaeta, 2001; Näätänen, Gaillard, & Mäntysalo, 1978; Polich, 2007; Verleger, 1988). The experiment consists of a listening task, in which a repeated sequence of tones is interrupted by sudden changes in pitch or loudness. In response to these 'oddball' tones, EEG activity shows characteristic changes shortly after presentation (mismatch negativity, P3a, and re-orientation negativity, see e.g. (Parmentier, 2014; Sussman, Winkler, & Schröger, 2003). Yet another, relatively late ERP is the N400, with a negative peak latency of approximately 400ms (Federmeier, 2007; Kutas & Hillyard, 1980). This component has been found to be a marker of semantic and lexical processing. Its magnitude is generally modulated by the degree of predictability of a word given the previous context (cloze probability; Taylor, 1953). This effect has been interpreted as a signature of lexical integration effort: if a word is less predictable given the previous semantic information, the brain performs more lexical search activity than in a highly predictable context (Obleser & Kotz, 2011).

More recent EEG as well as MEG (magnetoencephalography) studies, rather than using the traditional averaging technique of ERPs, have conducted experiments with continuous speech. This line of research has discovered that there is a measurable correlation between the ongoing auditory stimulus and neural brain oscillations in low-frequency regions. Specifically, cortical oscillations become synchronized with the broadband amplitude envelope of speech in delta/theta ranges (2-8Hz, e.g. Ahissar et al., 2001; Luo & Poeppel, 2007), with a latency of approximately 100-200ms (Baltzell et al., 2016). This so-called neural entrainment has been found to be higher to attended than to unattended speech in multi-talker listening situations (Ding & Simon, 2012; Kerlin, Shahin, & Miller, 2010; Zion Golumbic et al., 2013). Particularly for this reason, neural entrainment has been widely studied, both from a theoretical perspective and with a view to potential applications (e.g. 'hearables', cognitively steered hearing aids etc.).

Throughout this thesis, the terms 'temporal envelope', 'amplitude envelope' and 'envelope' are used interchangeably to refer to slow fluctuations in overall amplitude of speech signals. These convey various types of segmental cues of linguistic information, such as manner of articulation, voicing or vowel identity (Rosen, 1992). Specifically, temporal envelopes of speech were calculated by computing the absolute value of the Hilbert transform of the speech signals and filtering these with Butterworth filters between 0.1 and 30 Hertz.

Despite the growing research body on cortical entrainment, its exact working mechanism is still unknown. There are different hypotheses on the functional role of cortical entrainment, each with different implications for speech processing (e.g. Ding & Simon, 2014). Neural entrainment has been identified not just in humans, but also in animals (Kikuchi et al., 2017) and for non-speech sounds (Millman, Prendergast, Hymers, & Green, 2013; Steinschneider, Nourski, & Fishman, 2013) as well as reversed speech (Howard & Poeppel, 2010), indicating that cortical entrainment is partly an automatic response elicited by various kinds of auditory stimulation and not just intelligible speech. While higher entrainment and intelligibility tend to co-occur, the question remains which is a necessary precursor for the other: Is entrainment necessary for intelligibility, or does entrainment follow as a marker of intelligibility? (Zoefel & VanRullen, 2015a, 2015b) One recent study tested stimuli with bistable word repetitions (e.g. 'lamp' vs. 'plan' in French) and found that changes in the latency of entrainment, particularly in higher frequencies, was associated with the respective speech percept maintained by subjects (Kösem et al., 2018). While this mechanism might indicate that entrainment directly influences perception, the opposite position could also be taken. Thus, the exact relationship between entrainment and intelligibility is still unclear beyond the general agreement that these two phenomena tend to co-occur.

This thesis, rather than providing an examination of the currently disputed origins and mechanisms of cortical entrainment, operationalizes entrainment as the degree of synchronization of cortical oscillations with an auditory stimulus within a low-frequency range ('cortical tracking of speech'). It is measured through a correlation technique based on machine learning as implemented in the multivariate Temporal Response Function (mTRF) toolbox (Crosse, Di Liberto, Bednar, & Lalor, 2016). This correlation technique can be applied in two 'directions' of analysis: the 'forward model' and the 'backward model'. While the forward model predicts the neural response to the auditory stimulus and establishes the correlation between this predicted response to the actual response, the backward model reconstructs the auditory stimulus based on the multi-channel, time-varying neural response and then calculates correlations between this predicted and the actual auditory stimulus (Crosse, Di Liberto, Bednar, & Lalor, 2016). Unlike the forward model (the generation of predictions for the neural response as a function of the acoustic signal), the backward model does not require a pre-selection of specific neural response channels, as the model itself automatically assigns low weightings to channels that contribute less to the reconstruction than others (Mesgarani, David, Fritz, & Shamma, 2009). In addition, in contrast to the forward model, the backward model is able to account for inter-channel correlation in the data, as is often the case for EEG. This is possible because in the backward model, mapping is done from all channels simultaneously rather than from each channel individually (Mesgarani et al., 2009). For these reasons, this thesis used the backward model for all correlational mapping analyses of the frequency following (brainstem) and cortical entrainment.

Based on the successful application of computational models for cortical brain responses, the correlation technique described above has also been extended to other neural processing stages. The same signal reconstruction technique of stimulus-response mapping has recently been successfully applied to the early brainstem response (Etard, Kegler, Braiman, Forte, & Reichenbach, 2019; Forte, Etard, & Reichenbach, 2017; Maddox & Lee, 2018; Reichenbach, Braiman, Schiff, Hudspeth, & Reichenbach, 2016), as well as to the relatively late stage of lexical/semantic processing (Broderick, Anderson, Di Liberto, Crosse, & Lalor, 2018). Forte et al. (2017) used a competing speaker paradigm, similar to that for cortical entrainment and found that on average, the brainstem response to the fundamental waveform (a signal computed from the fundamental frequency, f0) was significantly higher to the attended than the unattended speaker. Broderick et al. (2018) conducted an experiment with different speech materials (competing speaker, audiovisual and time-reversed speech) and found that centro-parietal EEG channels only produced prominent negative responses when subjects understood the speech they heard, were paying attention or had additional visual cues when speech was presented in noise. This centro-parietal response with a time lag of approximately 200-600ms is strikingly similar to the ERP N400, which has been extensively studied in connection with semantic and lexical processing. These successful applications of the correlation technique to the brainstem and the lexical/semantic processing level are particularly interesting with a view to creating realistic listening environments for experimentation; whereas previously the brainstem response could only be measured to repeated syllables or vowels due to the small magnitude of the signal and the high levels of noise, the latest studies have been able to extract the brainstem response from continuous speech. Similarly, the response recorded in Broderick et al. (2018) did not require the careful construction of sentences with varying degrees of last-word predictability, but could be elicited using continuous speech only.

## 1.2 Decoding auditory attention

### 1.2.1 Bottom-up and top-down effects of auditory attention

Successful speech recognition in adverse listening conditions is not just simply a function of the type and degree of noise, but also requires attending to a speech stream while simultaneously filtering out irrelevant sounds. Much of the theoretical work on how humans direct auditory attention is inspired by research on visual attention

(Bronkhorst, 2015). The two domains share several characteristics: in both the visual and the auditory system, humans tend to group stimuli into perceptual 'objects' or streams, i.e. features that belong together (Pressnitzer, Sayles, Micheyl, & Winter, 2008; Shinn-Cunningham & Best, 2015). Two broad categories influence which object is in the listener's focus at any given time: bottom-up (e.g. physiological characteristics such as brightness, loudness) and top-down (e.g. the observer's goal) (Kaya & Elhilali, 2017). Both object formation (the grouping of features) and object selection (focus on a specific object) are important for successfully directing attention to a target stream (Ding & Simon, 2012; Griffiths & Warren, 2004; Shinn-Cunningham, 2008).

Regarding bottom-up factors, previous research has found that a variety of stimulus characteristics such as amplitude modulations (Arlinger & Gustafsson, 1991) or the energy allocation in the frequency spectrum of noise (Kaplan & Pickett, 1982; Prosser, Turrini, & Arslan, 1991) affect speech intelligibility in different ways. EEG studies have explored bottom-up, stimulus-dependent effects of auditory attention through ERP measures such as the P1-N1-P2 complex. Specifically the N1 component, a negative deflection of the EEG signal occurring approximately 100ms after stimulus onset and signalling the neural encoding of sound onset in auditory cortex, is described as an 'obligatory' or 'sensory' response (Steinschneider & Dunn, 2002). The P1-N1-P2 complex can be recorded without the participant's active doing; that is, the participant could watch a movie or read a book during the experiment (Martin et al., 2008). Such research indicates that at least certain early stages of the neural processing of sounds are automatic and driven by stimulus features.

In addition to such findings of the partly automatic processing of auditory stimuli, research has also indicated that the encoding and processing of stimulus features is malleable and can change with training and long-term experience: It has been demonstrated that, behaviourally, musically trained children and adults perform better on speech-in-noise tasks than non-musicians (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Strait, Parbery-Clark, Hittner, & Kraus, 2012), and that the more years people had spent practicing an instrument, the faster their reaction times were to a target stimulus in auditory experiments (Strait, Kraus, Parbery-Clark, & Ashley, 2010). This enhanced and faster performance on speech-in-noise task is also reflected on a neural level: adolescents with musical training were found to have enhanced response consistency on the brainstem level and, on the cortical level, a more mature P1-N1 complex response (Tierney, Krizman, Skoe, Johnston, & Kraus, 2013); children with musical training also exhibited less auditory brainstem response degradation with the addition of background noise compared to their non-musician peers (Strait et al., 2012).

There is also direct evidence that top-down factors shape the neural processing of attended and unattended speech. Research with neural entrainment in particular has demonstrated the immediate, online effects of selective attention to a speech stream at least on the cortical level. In quiet, there is no significant difference in cortical entrainment between an active and passive listening situation (Kong, Mullangi, & Ding, 2014). In a competing speaker situation, however, several studies have found that attention can selectively enhance the representation of the target speaker compared to the distractor (Ding & Simon, 2012; Kerlin et al., 2010; Zion Golumbic et al., 2013).

Computational models based on stimulus reconstruction have now become so accurate that in certain conditions, attention to a speech stream can be decoded on the basis of single-trial EEG data only (Fuglsang, Dau, & Hjortkjær, 2017; Horton, Srinivasan, & D'Zmura, 2014; O'Sullivan et al., 2015). In addition to top-down attention, contextual factors such as the presentation of congruent audiovisual stimuli have also been shown to enhance entrainment (Arnal, Wyart, & Giraud, 2011; Crosse, Butler, & Lalor, 2015), potentially because they assist in predicting the timing of incoming speech sounds. These top-down factors impacting entrainment thus provide evidence that, at least on the cortical level, attending to sounds and disregarding others is facilitated by an active, adaptive neural mechanism and not just driven by stimulus-dependent, bottom-up features.

While it is recognised that the top-down attentional selection of a specific stream is central for successful speech understanding, to what extent distractor streams are nevertheless processed in parallel to target streams has been a topic of debate (Holender, 1986). While early-filter theories (Broadbent, 1958) proposed that the attentional selection occurs prior to any semantic processing, late-filter theories suggested that distractor streams are at least partly processed up to a semantic level as well (e.g. Deutsch & Deutsch, 1963). Research with different types of maskers has found speech, compared to other maskers such as noise-vocoded speech or unmodulated noise, to be the most effective masker (Rosen et al., 2013), likely because there is interference on a semantic level. Further work with speech distractors in different languages supports these findings. Listeners displayed worse speech perception when the distractor is in a language they understand, compared to a language they do not (Lecumberri & Cooke, 2006; Rhebergen et al., 2005). Such findings support the proposition that distractor sounds, at least in certain conditions, are processed up to a high level.

While it is becoming clear that, on a neural level, attentional selection effects can be detected on a cortical level, there is also some evidence to suggest that the earliest effects of attention are measurable subcortically. Recent studies that have applied the correlation technique of cortical entrainment to the FFR have provided evidence that effects of selective attention can already be detected in the brainstem, with latencies as short as 10ms (Forte et al., 2017; Maddox & Lee, 2018; Reichenbach et al., 2016). Forte et al. (2017) used a competing speaker task similar to those employed for the detection of cortical entrainment. They instructed participants to either attend to a male or female speaker narrating audiobooks, and found that, on average, the response to the attended speaker was significantly larger than to the unattended, for both a female and a male speaker. These new results are particularly interesting because they are the first to give evidence of preferential encoding of speech streams at such a very early, subcortical stage.

### 1.2.2 Listening effort and its relationship with auditory attention

For certain situations and/or populations, it has been reported that paying attention to a target stream is experienced as more tiring and effortful. For example, people with hearing impairments consistently report higher subjective levels of effort in adverse conditions such as cocktail party situation (Ohlenforst et al., 2017). Importantly, however, this increased effort cannot easily be captured in behavioural measures such as speech recognition scores alone (Gosselin & Gagné, 2011). Although listening effort has been researched and theorized in considerable length and depth within the research community, there still remains fundamental disagreement with regards to how it should be conceptualized, whether it has merit in the first place and how it can be measured (McGarrigle et al., 2014). In many studies, listening effort is equated to or described similarly to cognitive load, a concept dating back to Kahneman (1973), who proposed that humans only have a certain limited capacity to process information at the same time. Cognitive load arises when the task(s) at hand exceed(s) that capacity. One recent, often-cited definition of listening effort comes from a white paper published by the British Society of Audiology (McGarrigle et al., 2014), which defined listening effort as the "mental exertion required to attend to, and understand, an auditory message". Compared to cognitive load, thus, listening effort applies specifically to auditory processing, and focuses more on the situational requirements for processing rather than the mental capacities of the listener.

Listening effort is particularly relevant in clinical applications such as the hearing aid industry, since people with hearing loss routinely report higher levels of effort and fatigue in noisy situations (Kramer, Kapteyn, Festen, & Kuik, 1997). Hearing impairment in general has been shown to impact quality of life in the form of greater difficulties with social participation, communication, as well as cognitive function and capacity, often resulting in social withdrawal and isolation (Heffernan, Coulson, Henshaw, Barry, & Ferguson, 2016; Solheim, Kværner, & Falkenberg, 2011). Additionally, people with hearing loss experience stigmatization due to a variety of negative connotations with the condition such as old age, incompetence, and cognitive and social impairment (Southall, Gagné, & Jennings, 2010; Wallhagen, 2009). Multiple studies have pointed out the negative impact hearing impairment – and the related higher listening effort – has on the psychosocial functioning of hearing-impaired people (Edwards, 2007), such as a greater need for recovery after work (Nachtegaal et al., 2009), higher incidences of sick leave due to hearing-related distress (Kramer, Kapteyn, & Houtgast, 2006) and higher rates of early health-related retirement than normal-hearing people (Danermark & Gellerstedt, 2004). Although listening effort alone is unlikely to be responsible for all these issues, there is an increased awareness of hearing-related problems that cannot be captured by speech audiometry alone (Winn, Edwards, & Litovsky, 2013).

While hearing-impaired people might find listening to speech in adverse conditions particularly effortful, they are not the only group for which listening effort has been shown to be relevant. Borghini & Hazan (2018) investigated listening effort in native (L1) versus non-native (L2) listeners using pupillometry and a speech-in-noise task. They adjusted noise levels individually to equate intelligibility levels across all listeners, and still reported increased effort in the L2 listeners, compared to L1 listeners. In other words, despite equal levels of behavioural performance, L2 listeners need to invest greater resources to understand speech in noise. Although part of the reason for this increased effort (in the sense of greater resource investment) might be shared between the two groups, there are also several factors that differentiate the two: for example, while hearing impairment is associated with peripheral degradation resulting in reduced audibility of the speech, L2 listeners do not suffer from such bottom-up issues. In contrast, L2 listeners might not be fully proficient in the target language, introducing another set of difficulties such as interference of the native language on a lexical level (Titone, Libben, Mercier, Whitford, & Pivneva, 2011), whereas hearing-impaired listeners do not have these difficulties.

In normal-hearing, native listeners, effort has also been shown to vary, particularly as a function of task demand. Multiple studies have used a dual-task paradigm, where subjects are instructed to carry out a secondary task (for example a visual search paradigm, or a memory task) in addition to the primary listening task. A recent review of 29 studies using the dual-task paradigm (Gagné, Besser, & Lemke, 2017) has concluded that most studies have successfully found differences in the effort expended between easier and more difficult listening tasks.

While listening effort is a useful and important concept to consider in many contexts, this thesis does not exclusively focus on the investigation of listening effort for the following reason: listening effort is closely linked to and influenced by factors like intelligibility, attention, and motivation, which themselves are hard to pin down with a single measure. Listening effort is difficult to measure without a preconceived operationalization of what exactly it constitutes: for example, if pupil dilation is taken as the primary indicator of listening effort, any changes in pupil size over time will be attributed to a change in listening effort. While this approach is worthwhile in its own right, this thesis purposely does not make a claim to measure listening effort in a direct way. Rather, with the experiments and analyses presented here, the aim was to gain analytical insight about listening effort indirectly through the exploration of different measures. It is through this joint analysis of different measures and conditions that I hope to draw conclusions about the steering and manipulation of auditory attention, and indirectly listening effort, at different stages of neural language processing.

### 1.2.3 Measures to investigate auditory attention and language processing

There is a variety of measures applied to investigate speech processing, attention and effort. Studies generally report one (or a combination) of the following categories of measures: (1) self-report (subjective measures): usually questionnaires, also often used in combination with other measures, (2) behavioural measures: single-task (e.g. button press, response rate etc.) vs. dual-task paradigm (measure of attention allocation, modelled after Kahneman's cognitive load theory), and (3) neurophysiological measures: fMRI (functional magnetic resonance imaging) (Mulert et al., 2008), EEG (Bernarding, Strauss, Hannemann, & Corona-Strauss, 2012; Song & Iverson, 2018), pupillometry (Borghini & Hazan, 2018; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017) and skin conductance (Holube, Haeder, Imbery, & Weber, 2016). Each of these three categories has its own merits and advantages, but also specific challenges. While the multitude of approaches, paradigms and techniques available on the one hand has led to an abundance of research with different perspectives on the issue, it on the other hand makes comparisons across different studies difficult. Unfortunately, studies investigating more than one measure in the same experiment are rare (e.g. Miles et al., 2017; Mulert et al., 2008), which makes it difficult to compare different methodologies with one another and tease apart the specific mechanisms they investigate.

Subjective measures, particularly those embedded in the field of listening effort, are by some practitioners regarded as the gold standard when it comes to clinical applications such as hearing aids. Particularly in the hearing aid industry, where hearing aid discontinuation of use is known to be a widespread issue, it has been argued that it is the subjective feeling about exertion, hearing comfort and benefit that matters most in the end. However, it is not always clear whether subjects in experiments rate the same dimensions when asked about listening effort (e.g. fatigue, level of performance, motivation), and ratings are difficult to compare directly between subjects, as for some an 'average' rating might lie higher on a scale than for others. Moreover, subjective ratings are crucially dependent on the phrasing of the question, for which there is no agreed-upon standard in listening effort research (e.g. NASA task load index, ACALES, Hart, 2006; Krueger, Schulte, Brand, & Holube, 2017). It is also known that subjective ratings can vary as a function of several factors such as motivation to perform the task (Picou & Ricketts, 2014).

Behavioural measures of auditory attention encompass a wide variety of different tasks, ranging from word recognition scores to search paradigms. In auditory tasks, word recognition scores often serve as a proxy of intelligibility, and are thus widely used. The advantage of behavioural tasks, compared to subjective tasks, is that they are a more objective measure of performance in challenging conditions. Yet, behavioural measures are influenced by a number of variables, such as the task at hand, as well as 'external factors' such as the subjects' motivation. Particularly in dual task paradigms, a careful choice of experimental design is essential in order to distinguish between different levels and types of load, as these have been shown to affect distractor processing and interference in contrasting ways (Lavie, Hirst, De Fockert, & Viding, 2004).

The main advantage of objective measures is that they entirely circumvent many of the problems associated with subjective and behavioural measures (e.g. deciding on a particular task to be carried out in experiment, or a particular question in a questionnaire), and can be recorded without the subject's active 'doing'. Nevertheless, even these measures have their challenges. Since the human brain is highly complex, it is not always trivial to determine which factors the change in the respective measure under investigation (e.g. blood flow in MRI studies, neural spiking in EEG and MEG studies) can be attributed to.

### 1.2.4 Hearing impairment and its effects on auditory attention and speech processing

While a better understanding of neural language processing in difficult circumstances can benefit a wide variety of fields, such as the development of speech recognition machines, or the improved acoustic design of public spaces (such as train stations), there is one particular group that would profit the most: the hearing-impaired. Hearing impairment, particularly age-related sensorineural hearing loss, is a growing issue and one of the most common disabilities all over the world. According to the WHO, 466 millions of people worldwide (over 5%) currently suffer from disabling hearing loss (WHO, 2018). Due to ageing societies and other factors, this number is expected to rise to 900 million by 2050, amounting to about 10% of the world's total population. Despite the fact that hearing loss is so widespread, many people that could profit from a hearing aid do not use one. Statistics vary, but the percentage of people who could benefit from a hearing aid but have never used one range from 60% in the UK (Action on Hearing Loss, 2018) up to 84% in the US (Hearing Loss Association of America, 2018). In other words, even in developed countries, far less than half of all the people that could benefit from hearing aids have ever used them. There is a variety of reasons for the fact that the hearing aid market remains drastically underserved: in developing countries, many people cannot afford to buy a hearing aid or do not have access to audiological services. In developed countries, even people who have bought a hearing aid do not always use it consistently: It is reported that, for a variety of reasons, up to 40% of hearing aid owners either entirely discontinue to use their hearing aids within months of buying it or never gain optimal performance (Barker, Mackenzie, Elliott, Jones, & de Lusignan, 2016). Among other issues such as difficulty in handling and maintenance (Hartley, Rochtchina, Newall, Golding, & Mitchell, 2010; Kochkin, 2000), many users report continuing dissatisfactory hearing performance, particularly in social environments such as bars, cafes or restaurants (Cord, Surr, Walden, & Olson, 2002). Aware of the problem at hand, hearing aid manufacturers have been seeking ways to mitigate these issues and particularly aid with speech perception in 'cocktail party' situations.

One of the main complaints of hearing aid users is the unsatisfactory situation in noisy environments such as cafes and restaurants (Pichora-Fuller et al., 2016), where it is hard for them to pay attention to a specific speaker, while ignoring background noises or other voices. Previous research supports these subjective reports and provides evidence that hearing-impaired people are at a disadvantage with regards to speech perception in adverse conditions, compared to normal-hearing people (Hagerman, 1984; Hopkins, Moore, & Stone, 2008; Plomp, 1986)

Speech perception for hearing-impaired listeners is different from non-impaired speech perception on several levels. At the peripheral level, damage to outer hair cells in the cochlea affects how well the incoming signal is transmitted to the auditory nerve. This damage results in the loss of the active mechanism, through which amplitude and frequency selectivity of incoming sounds are enhanced in a nonlinear way in the healthy ear (Ashmore, 2008). Due to the damage and the loss of the active mechanism, hearing-impaired people suffer from decreased frequency selectivity, that is the ability to identify and distinguish tonal components in complex sounds (Dubno & Schaefer, 1992). In addition to less precise frequency selectivity, the active mechanical response in the cochlea which amplifies low-level sounds and compresses high-level sounds is impaired (Robles & Ruggero, 2001). This means that, on the one end of the spectrum, the amplitude of weak sounds is less well enhanced at the basilar membrane than it would be with an intact active mechanism. On the other extreme, hearing-impaired people are more sensitive to loud sounds, as the active compression mechanism functions less well than in normal-hearing people (Moore, 1995; Moore, Glasberg, & Vickers, 1999).

These physiological changes and the damage to outer hair cells in the cochlear leads to increased difficulties for speech perception in hearing-impaired listeners. While normal-hearing people get significant benefits from ‘glimpsing’ the target during temporal dips of fluctuating maskers (Cooke, 2006), the hearing-impaired profit significantly less (Moore, Peters, & Stone, 1999), due to the poorer resolution of the auditory signal. Listeners with hearing loss also perform more poorly at discriminating the fundamental frequency ($f_0$) of complex sounds (Bernstein & Oxenham, 2006) and cannot use the temporal fine structure cues of speech as well as normal-hearing people (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006).

Fine structure cues will be defined as the "variations of wave shape within single periods of periodic sounds, or over short time intervals of aperiodic ones" (Rosen, 1992). These, in contrast to the slow fluctuation rates of the temporal (broadband) envelope of speech, are characterized by rapid oscillation rates from about 600Hz up to 10kHz. The auditory system performs a spectral analysis of incoming sounds with an array of overlapping auditory filters, each centering on a particular frequency band. The output of each filter, in turn, is like a band-pass filtered signal characterized by its temporal envelope and temporal fine structure (Lorenzi et al., 2006; Moon & Hong, 2014; Moore, 2008).

As mentioned above, studies have demonstrated that hearing-impaired people have a reduced or no ability to use temporal fine structure cues, particularly – but not exclusively – at higher centre frequencies (Hopkins & Moore, 2007; Lacher-Fougère & Demany, 2005; Lorenzi et al., 2006; Moore, 2008; Santurette & Dau, 2007). At the same time, there is behavioural evidence to suggest that hearing-impaired people are better at detecting envelope modulations than normal-hearing people (Ernst & Moore, 2012; Moore, Wojtczak, & Vickers, 1996; Sek et al., 2015; Wallaert, Moore, Ewert, & Lorenzi, 2017). Since it has been demonstrated that, in normal-hearing people, perceptually enhanced envelope modulations are associated with worse speech perception (Moore & Glasberg, 1993), it could be concluded that poorer speech perception in hearing-impaired people is related to a processing imbalance between envelope and temporal fine structure cues.

In addition to bottom-up, sensory-related deficits, there is evidence to suggest that hearing impairment affects auditory attention and the attentional selection of sound streams (Dai, Best, & Shinn-Cunningham, 2018). In order to support hearing-impaired people in challenging situations, hearing aids nowadays commonly have a range of programs and algorithms for speech processing built in. These algorithms are designed to amplify sounds from certain directions or with certain characteristics to aid speech in noise perception. However, hearing aid users still do not take full advantage of these, mostly because the range of algorithms available currently is still unable to provide satisfactory online real-time processing (Cord et al., 2002; Ricketts, Picou, & Galster, 2017). This is a highly complex problem and research departments in hearing aid companies worldwide are working on improved solutions. One of the directions in which manufacturers conduct research is on auditory attention, with the aim of building hearing aids that can detect what the user is trying to listen to, and amplify that specific speech stream ('cognitively controlled hearing aids'). There is some research that has been delivering first promising results, such as pilot experiments with eye-gaze steering for selective auditory amplification (Favre-Felix, Graversen, Dau, & Lunner, 2017), and beamforming algorithms that can be steered adaptively based on auditory attention decoding through EEG (Das, Van Eyndhoven, Francart, & Bertrand, 2016). However, because the selective attention problem and appropriate support for hearing-impaired people is such a complex challenge, to date there are still no products on the market that can actually provide compensation for hearing loss in a similarly effective way as glasses do for loss of eyesight.

With the rise of technologies such as machine learning, and these being applied to a wide variety of fields, spanning from speech recognition to the hearing aid industry, natural language processing is advancing fast. Speech enhancement techniques apply different kinds of processing to the signal in order to make the target better understandable (i.e. improve intelligibility) or to increase listening comfort (Bentler, Wu, Kettel, & Hurtig, 2008). Despite many remaining hurdles relating to functionality and performance, speech enhancement techniques have had great success and can be applied in a wide variety of settings such as telephone conversations, headsets (noise cancelling function) and hearing aids. This thesis specifically investigated two of the most commonly applied algorithms in hearing aids: noise cancelling and directional microphones (beam forming) (Desjardins & Doherty, 2014). While these signal processing techniques have been shown to provide benefits for listeners in laboratory conditions, much less is known about their real world use for hearing-impaired populations (Bentler, 2005).

Directional microphones operate through increased sensitivity to sounds from certain directions (typically the front) compared to others (typically the side or back) (Ricketts et al., 2017). Various studies under laboratory conditions have demonstrated increased speech recognition performance in noise with the directional microphone mode, compared to the omnidirectional mode (Picou, Aspell, & Ricketts, 2014; Ricketts &

Hornsby, 2006). However, one study that investigated both laboratory and real world conditions found that hearing aid users do not report significant performance increases when using the directional microphone mode (Walden, Surr, Cord, Edwards, & Olson, 2000). In fact, when fitted with manual switching between directional and omnidirectional mode, a substantial percentage of both adults and children did not consistently switch to the optimal program for a given SNR (Cord et al., 2002; Ricketts et al., 2017).

The principle of digital noise reduction is to reduce the gain of background noise and preserve or enhance that of incoming speech in the hearing aid. The implementation of this principle varies widely across systems, with regards to the amount of gain reduction, SNR thresholds for activation, and speed of algorithm engagement (Bentler & Chiou, 2006). A number of studies has reported no improvement in speech recognition scores for noise reduction algorithms (Alcántara, Moore, Kühnel, & Launer, 2003; Boymans & Dreschler, 2000; Brons, Houben, & Dreschler, 2015; Desjardins & Doherty, 2014; Ricketts & Hornsby, 2005). However, digital noise reduction features have been shown to improve other measures such as sound quality (Ricketts & Hornsby, 2005). A few studies have also reported decreased listening effort with NR algorithms as measured through a dual-task design (Desjardins & Doherty, 2014) or pupillometry (Wendt, Hietkamp, & Lunner, 2017). It might well be that the failure to demonstrate benefits in terms of speech recognition scores is due to ceiling effects: the in studies commonly used measure of speech reception threshold (SRT) in people with mild to moderate hearing impairment corresponds to SNRs of between -10dB and 0dB. NR algorithms, however, have been shown to be most effective in positive SNRs (Fredelake, Holube, Schlueter, & Hansen, 2012; Smeds, Wolters, & Rung, 2015), which are in turn by far the most common signal to noise ratios listeners are exposed to in real life (Buchholz et al., 2016; Smeds et al., 2015).

Overall, despite the fact that both directional microphones and noise reduction features have been implemented in hearing aids for decades, it has been notoriously difficult to prove their benefit to the wearer with an objective measure. This challenge is partly driven by the fact that the frequently reported 'higher effort' as a subjective measure is difficult to quantify, and has been operationalized by studies in different ways. Particularly in the case of noise reduction, speech recognition scores do not appear to be satisfactory measures of the perceived and actual benefits to the user. More objective, neural measures of speech processing as developed in this thesis could provide helpful insights into the origins and working mechanisms of the advantages that digital signal processing features implemented in hearing aids yield.

### 1.2.5 The role of motivation as a top-down factor in neural language processing

While it is clear from previous research that bottom-up factors such as the type and level of masking have a considerable influence on speech perception in challenging situations, it has also been postulated that top-down, cognitive factors affect speech recognition in general, and selective auditory attention in particular (Shinn-Cunningham, 2008). The manipulation of top-down factors such as motivation to complete a difficult listening task can shed light on the extent to which neurophysiological measures in particular are shaped not just by stimulus-dependent, bottom-up factors, but also by cognitively controlled top-down mechanisms.

It can be assumed that the attentional resources needed to complete a task have an inverted U-shaped form, as a function of task demand. For very easy tasks, it is not necessary to allocate many attentional resources in order to succeed. With increasing task demand, however, attentional resources presumably need to rise accordingly. At some point, when the task difficulty exceeds the cognitive capacities of the person, attention can be expected to fall again. Cortical entrainment, as a neural measure of selective attention, might reflect certain aspects of this inverted U-shaped relationship between attentional resource allocation and task difficulty. In simple single-talker situations, it has been found that neural entrainment to speech is not enhanced, even when attention is allocated to that speaker (Kong et al., 2014). However, when a competing speaker is added to the scene, and subjects are instructed to only pay attention to one speaker, cortical tracking of the attended speaker shows a robust advantage compared to the ignored speaker (Ding & Simon, 2012; Kerlin et al., 2010; Zion Golumbic et al., 2013). Finally, in the case when paying attention becomes impossibly hard due to a low signal-to-noise ratio, entrainment eventually drops away again (Ding & Simon, 2013).

Previous research suggests that motivation has a similar task-dependent shape: psychological studies show that the motivation to achieve a goal is at its highest when the task is challenging, but not so difficult that it is perceived as impossible. This so-called 'Yerkes-Dodson law' was first demonstrated on rats that had to find their way out of a maze (Yerkes & Dodson, 1908). When given mild electrical shocks, the rats were more likely to complete the task than without any shocks, whereas strong electrical shocks caused the learning performance to drop. Although this observed pattern was much less consistent when tested with baby chickens (Cole, 1911) or kittens (Dodson, 1915) in two follow-up studies with the same experimental design, many studies have built on this paradigm and extended its implications about the U-shaped curve of performance to other domains such as work-related stress (for a review, see Corbett, 2015).

However, little research has been conducted so far on the role of motivation for neural speech processing and attention. One study that examined the effect of motivation on subjective listening effort operationalized motivation as either the presence or absence of a quiz at the end of a listening task that subjects had to complete. It was assumed that subjects were more motivated to listen carefully when they knew that they would later be asked questions about it. Results indicated that motivation generally increased listeners' subjective ratings of listening effort and tiredness (Picou & Ricketts, 2014). Apart from this study using subjective and behavioural measures only, not much research has been conducted on the effect of motivation on language processing. While there is a body of research examining listening effort with neurophysiological measures such as pupillometry, EEG or skin conductance, these studies rarely mention motivation as a factor, and if so, not as one that is explicitly manipulated in the experimental design.

Yet, the effect of motivation manipulation on neurophysiological measures of auditory attention and language processing is interesting to study for the following reasons: (1) manipulating motivation allows us to potentially influence attention without changing the actual stimulus (which is usually done through some kind of signal degradation) (2) demonstrating that motivation influences neural measures of language processing would be another indicator that these neural measures are not just influenced by bottom-up, stimulus-related factors, but are also impacted by top-down cognitive processes and mental states. While this has been shown with attention decoding and neural entrainment, no other top-down measure has yet been demonstrated to impact neural measures of language processing, and (3) related to this point, investigating another top-down factor (motivation) allows us to more broadly examine the interplay between bottom-up and top-down effects in neural language processing.

### 1.2.6 The current thesis

This thesis investigates speech perception in a variety of difficult listening situations with different listener groups. The focus is auditory attention and speech comprehension as a multi-stage, multi-dimensional phenomenon, using a toolbox of subjective, behavioural and neurophysiological measures. The studies conducted for this thesis aimed to shed light on different aspects of auditory attention and speech processing in adverse conditions, with the common goal of better understanding differences between degradation types, levels and various listener groups. Specifically, this thesis addresses the following questions:

(1) What measures can be used to investigate speech processing and auditory attention in difficult listening situations? How are neurophysiological measures of speech processing affected by different degradation types and degrees, and how do the measures differ from one another? How does the joint analysis of several measures impact our understanding of neural language processing and auditory attention in adverse conditions?

(2) How do hearing aid users differ from normal-hearing listeners with regards to online markers of neural speech processing? How do processing algorithms commonly implemented in hearing aids affect neural measures of speech processing?

(3) What is the role of motivation as a top-down factor for listening effort and auditory attention modulation? Are there measurable differences in neural speech processing as a function of motivational manipulations? Can these manipulations be informative about the impact of top-down factors on neural speech processing more generally?

The first study of this thesis (Chapter 2), served two concurrent purposes: first, to establish a toolbox of neurophysiological measures to investigate auditory attention and speech processing; and second, to apply this toolbox to an experiment with different types and levels of maskers. A dichotic listening paradigm was used in order to examine the effects of signal degradation on the target in the presence of an unmanipulated distractor speaker. The target speech was degraded with different maskers. With the range of measures extracted from the EEG signal, neural language processing and auditory attention were compared at different stages, and the effect that different forms and levels of degradation have on these.

In the second EEG experiment (Chapter 3), measures from the toolbox developed in Study 1 were applied to investigate the effect of speech enhancement algorithms commonly applied in hearing aids (noise cancelling and beam forming/directional microphone). A hearing-impaired group as well as a normal-hearing control group were tested on a speech-in-noise task. Prior to the EEG recording, participants were fitted with hearing aids (hearing-impaired people for hearing loss compensation, and normal-hearing people with minimal amplification in order to ensure comparability between the groups). In addition to testing the impact of speech enhancement algorithms on neural processing, a subjective measure of listening effort was also collected in both listener groups.

The third study of this thesis (Chapter 4) details an EEG study that investigated the effect of motivation as a top-down factor for auditory attention, as well as a specific type of signal degradation and its effect on cortical entrainment. A group of normal-hearing subjects were tested with the battery of measures developed in this project, and stimuli similar to that of Study 1 were used for comparison. Specifically, this study was designed to make the link between effort, attention and various neurophysiological measures more explicit. In parallel, a hypothesis on the reasons for increased entrainment found in older, hearing-impaired people was investigated.

# Chapter 2: The effect of different maskers on neural speech processing

## 2.1 Introduction

How well speech is perceived in noisy conditions depends on the interplay of both bottom-up factors such as the acoustic characteristics of the noise, and top-down central and cognitive factors such as the ability to separate sound streams and pay attention to a specific one (Kaya & Elhilali, 2017). While previous research has investigated many different specific aspects of speech comprehension in adverse conditions, it is not well understood how various processing stages in the brain are affected differently by the interplay of bottom-up and top-down effects of degradation and attention. In contrast to behavioural measures, neural markers of speech processing offer a direct perspective on the working mechanisms of different processing stages beyond the verbally/behaviourally measurable. In particular, research on the brainstem (Forte et al., 2017) and the cortical level (Ding & Simon, 2012; O'Sullivan et al., 2015) has developed tools to measure neural processing of continuous speech, which can be applied to a variety of adverse listening conditions.

This study examined neural measures of speech processing in conjunction with a behavioural measure of speech recognition of a target speaker, in the presence of a distractor speaker as well as added degradation (3 levels of noise and vocoding). The research focus was on how various stages of neural speech processing of the target are affected by degradation in different ways. While previous research has mostly focused on varying distractors and investigating the impact of these manipulations on target processing (e.g. Lecumberri & Cooke, 2006; Rhebergen et al., 2005; Rosen et al., 2013), this study investigated the intelligibility and form of the target speech without modifying the distractor. A toolbox of EEG measures (FFR, cortical entrainment, N400) was used to extract neural speech processing at different stages in degraded conditions.

Previous research has shown that the auditory brainstem response is highly sensitive to signal degradation. When speech is degraded in noise, the auditory evoked responses of the brainstem become delayed and reduced (Anderson, Skoe, Chandrasekaran, & Kraus, 2010). Several studies have investigated the effect of noise degradation on the FFR and found that while onset portions of the brainstem response are affected the most, even low levels of noise (SNRs of +5 to +10dB) adversely impact the precision of amplitude and phase tracking of the FFR, which in turn is connected to worse speech perception performance (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2013; Burkard & Sims, 2002; Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; Musacchia et al., 2018; Russo, Nicol, Musacchia, & Kraus, 2004).

To this date, it is unclear to what extent top-down processes such as attention modulate the FFR, or whether this early response is an entirely automatic process, driven by acoustic features of the stimulus only. While some studies have found the brainstem response to vary with attention (Galbraith & Arroyo, 1993; Galbraith, Bhuta, Choate, Kitahara, & Mullen, 1998; Galbraith & Doan, 1995; Lehmann & Schönwiesner, 2014), others are more inconclusive or indeed present evidence to the contrary (Holmes, Purcell, Carlyon, Gockel, & Johnsrude, 2018; Varghese, Bharadwaj, & Shinn-Cunningham, 2015). Some of this variation in findings may be attributable to methodological challenges: many studies have used simple tones, single vowels or syllables, rather than continuous speech. Since the brainstem response is very small, the traditional approach requires many repetitions of the stimulus, which in turn may make it difficult for subjects to sustain attention. In contrast, using continuous speech as a stimulus could be more informative particularly for auditory attention research. Measuring FFR to speech - as opposed to clicks or vowel sounds with many repetitions - is a recent development made possible to the signal reconstruction techniques used in this thesis (Forte et al., 2017; Maddox & Lee, 2018; Reichenbach et al., 2016). One such study has showed that the brainstem response is larger when speech is attended (Forte et al., 2017).

In contrast to the brainstem response, the effects of top-down attention effects on cortical entrainment are better known. Paying attention to a speaker while ignoring another one has been shown to increase entrainment only for the target, but not the distractor (Ding & Simon, 2012; Kerlin et al., 2010; Zion Golumbic et al., 2013). Concerning the impact of bottom-up effects caused by different types and degrees of masking, cortical entrainment appears to be relatively robust to degradation in noise: Ding & Simon (2013) have shown that entrainment remains relatively unaffected by background noise apart from the most degraded condition measured (-3dB SNR). However, removing some of the temporal fine structure through vocoding on top of the noise significantly impairs the neural representation (Ding, Chatterjee, & Simon, 2014). This might indicate that noise (as a degradation added to the signal) and vocoding (as the direct degradation of the stimulus itself) are processed in different ways in auditory cortex.

The N400, an ERP component associated with high-level processing of lexical/semantic information, has been central to the debate discussing the point that top-down attention effects impact neural processing (early vs. late attentional selection theories) (Kutas & Federmeier, 2011). However, experiments have shown the N400 to display characteristics of both and thus not fitting neatly into either category. One study has found that the expectation of meaningful speech increased the N400 of syllables and made it more similar to the N400 measured in response to whole words (Bonte, Parviainen, Hytönen, & Salmelin, 2006), indicating that top-down processes such as expectations affect the N400. Other, related research has demonstrated that a greater N400 can indicate more effortful semantic integration (Chwilla, Brown, & Hagoort, 1995) and that the N400 is correlated with measures of listening effort such as pupil dilation (Kuipers & Thierry, 2011). A recent study (Song & Iverson, 2018) that measured both cortical entrainment as well as the N400 effect in an experiment with native and L2 listeners found that L2 listeners had greater lexical processing for high-predictability sentences than native listeners in their native language. Furthermore, when native listeners heard L2-accented speech, they also displayed greater N400 responses, suggesting that lexical processing of this speech was more effortful. Beyond demonstrating that the N400 is impacted by selective attention effects, all these studies also highlight the close link between attention and effort, and the difficulties of teasing the two apart.

This study explored different measures of neural language processing from peripheral to central units, and the effects that different types and levels of masking have on these measures. Subjects performed a two-talker listening experiment, and were instructed to attend to one speaker only. The target speaker was presented with different masker types and levels (noise and vocoding). The stimuli material consisted of three types of sentences: high cloze probability, low cloze probability and anomalous sentences. Cloze probability is here defined as the probability of a target word completing a particular sentence (Taylor, 1953). Subjects were instructed to press a button when they heard an anomalous sentence. This button press was collected as behavioural data and served as a measure of intelligibility. From the EEG recording, brainstem and cortical entrainment to both target and distractor speaker were extracted, as well as N400 for high and low cloze probability sentences. The following questions were investigated in this study: How is the neural representation and intelligibility of a target speaker affected by different levels and types of degradation? Does this vary across different levels of neural processing? How is an intelligible, non-degraded distractor speaker processed in such adverse listening conditions?

## 2.2 Materials and methods

### 2.2.1 Subjects

Twenty-two normal-hearing adults (10 = female, 12 = male, mean age = 32.5 years) participated in the experiment. All participants were native Standard Southern British English speakers. Subjects were paid, and the experimental procedure was approved by the UCL institutional ethics committee. Written informed consent was obtained before the experiment.

### 2.2.2 Stimuli

Sentence materials constructed previously at the Speech, Hearing and Phonetic Sciences, UCL (Stringer, 2015) were used. These were 960 sentences spoken by female and male British English speakers. The materials consisted of three different sentence types:

a) high cloze probability (e.g. "Doctors try to cure dangerous *diseases*.")
b) low cloze probability (e.g. "Scientists try hard to stop different *diseases*.")
c) semantically anomalous (e.g. "Doctors try to cure dangerous *pianos*.")

Speakers were presented in different ears (female target in the right ear, male distractor in the left ear). The type and amount of degradation was varied in the target speaker channel only. Intelligibility levels were chosen following a behavioural listening experiment with 10 participants. There were five different target stimuli types:

a) normal: no processing on the target speaker
b) high SNR: SNR +3dB (speech shaped noise matched to target speaker)
c) medium SNR: SNR -0.5dB
d) low SNR: SNR -4dB
e) vocoding: 14 channels, noise vocoded target speech

The distractor speaker was always presented without any processing. Target and distractor speech were matched in duration, so that onset and offset of both speech streams occurred at the same time.

### 2.2.3 Apparatus

An EEG experiment was conducted (Biosemi Active Two system with 64 (Ag/AgCl) electrodes) with 7 reference electrodes placed on the face to control for eye blinks and movement as well as muscle activation-induced artefacts. Participants were seated in a sound-attenuated booth. The experimental task took approximately one hour to complete.

### 2.2.4 Procedure

Participants were presented with a dichotic listening task and instructed to attend only the female speaker while disregarding the male speaker. Stimuli pairs were randomized for each participant and presented in eight separate blocks to the participant. Each block took about five minutes to complete, and after every second block listeners were given a short break. Before the main task, participants completed a trial run to familiarise them with the task. Subjects were instructed to press a button if they heard a semantically anomalous sentence (14.8% of all trials). This task served as a behavioural measure of speech intelligibility and ensured that participants were indeed paying attention to the task and the target speaker.

### 2.2.5 EEG analyses

Data recorded with the BioSemi Active Two system were analysed in MATLAB (version R2016b; MathWorks, Natick, MA) after being converted into MATLAB format with the function pop_biosig from EEGLab (Delorme & Makeig, 2004). Data were recorded at 8,192 Hz. Two electrodes placed on the mastoids were used as reference. Noisy channels were interpolated. For brainstem analyses, a high-pass filter of 0.1 Hz was applied and data downsampled to 4,096 Hz. For cortical analyses, data were band-pass filtered between 0.1 and 30 Hz and downsampled to 256 Hz. All filters used were zero-phase causal Butterworth filters of the 3$^{rd}$ order (response slope of -18dB per octave) as implemented in the ERPlab toolbox (Lopez-Calderon & Luck, 2014) of EEGlab (Delorme & Makeig, 2004). Artefact rejection was performed automatically on all trials, with a rejection threshold of ±150 μV. Independent Component Analysis was applied manually and components associated with eye movements as well as eye blinks with highly spatially stereotyped scalp projections (Onton & Makeig, 2006) were removed from the data. All pre-processing procedures were performed in Matlab using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011).

**FFR.** An analytical fundamental waveform was computed from the speech materials through the following method: first, the signal was filtered to extract the first harmonic. This was done by measuring the fundamental frequency (f0) contour using the program ‘fxrapt’ as implemented in Matlab (Talkin, 1995), dividing the signal into 250ms windows, and using a zero-phase low-pass Butterworth filter of the 5$^{th}$ order in each segment set to 50% greater than the maximum f0 in the window. The envelope of this first harmonic was remodulated so that it matched the low-pass broadband envelope of the signal. This was done by calculating the Hilbert envelopes of the low-pass filtered (50 Hz) original signal, and adjusting the envelope of the first harmonic to match these.

To the EEG data, a Butterworth filter between 100 and 300 Hz was applied to concentrate on frequency regions centering on the FFR. The mTRF toolbox (Crosse et al., 2016) in MATLAB was used to fit multivariate Temporal Response Functions relating the EEG data for each subject to the analytical waveform computed as described above. This toolbox enables the linear mapping between a stimulus and a response, and the response function resulting from this mapping can be used to predict a new stimulus or response (depending on whether the forward or the backward model is used). For the present thesis, the backward model was used in all analyses. This means that after training the model on a set of data (i.e. target and distractor sentences and their matching EEG response data), the spectrotemporal information of a previously unseen stimulus (sentence) was reconstructed given its matching response (EEG) data. Entrainment results were then computed as the similarity of the reconstructed stimulus and the actual stimulus. The mTRF toolbox uses the machine learning technique of ridge regression, a form of multiple regression that can account for data suffering from multicollinearity. This is the case for EEG data, as single electrodes are often highly linearly related with one another.

Specifically, the following procedure was applied to analyse the FFR: after matching the respective EEG data with each audio sample (i.e. with the analytical fundamental waveform of each sentence), the mTRF model for both target and distractor was trained by performing a (ridge) regression analysis between the stimulus and the response data, using time lags of -15ms to 45ms in order to avoid edge effects impacting crucial FFR latencies. Through this analysis, a set of channel weights over time were generated that linearly mapped the EEG responses back to the original auditory signal. These model weights indicate which latencies and electrodes carry the most information with regards to the matching of stimulus and response. After the training phase, the model predicted previously unseen data based on the response function generated. This means that the model reconstructed the fundamental waveform of each sentence (test set), based on the EEG data and the corresponding fundamental waveforms of the other sentences (training set). As a final step, correlation values between the reconstructed and the actual fundamental waveforms were computed for each condition separately. These correlations are indicative of the level of FFR tracking in each condition.

**Cortical entrainment.** For the purposes of this thesis, cortical entrainment was operationalized as the degree of synchronization between cortical oscillations (measured through EEG) and the temporal (amplitude) envelope of speech (‘cortical tracking of speech’). The amplitude envelopes of all sentences were computed prior to coherence analysis; the speech signals were full-wave rectified and filtered using the same high-pass (cut-off: 0.1 Hz) and low-pass Butterworth filters (cut-off: 30 Hz) that were used for the EEG signals. To match the sampling rate of the EEG data, audio data were then downsampled to 256Hz. To calculate cortical entrainment, the same procedure as for FFR tracking analysis was applied: the EEG data were first matched with the respective amplitude envelope of each sentence. Then, the backward model was trained for target and distractor, using time lags from -150ms to 500ms. The predictions generated from the model (i.e. the reconstructed amplitude envelopes) were then compared with the actual amplitude envelopes. However, rather than using correlation as a measure of similarity between these two signals, for cortical entrainment coherence was used to display entrainment in the frequency domain. This was done by segmenting the data in to 1-sec Hann windows with 50% overlap, and calculating coherence from the cross-spectral density of the FFT of the two signals, divided by the power spectrum of each signal.

**Lexical processing.** Data were downsampled to 256 Hz. The data was segmented into epochs time-locked to the final-word onsets (200ms pre-stimulus and 800ms post-stimulus intervals). Trials with amplitude exceeding ±150 μV were rejected, and the rejection rate averaged across subjects was 4.5%. N400 for the target speech was calculated separately for each condition by averaging the response to high vs. low predictability sentences respectively. The main statistical analysis was then performed using a narrower time window (300-500ms) and smaller midline electrode set (Fz, FCz, Cz, CPz, and Pz), following previous N400 studies (Song & Iverson, 2018; Strauß, Kotz, & Obleser, 2013).

## 2.3 Results

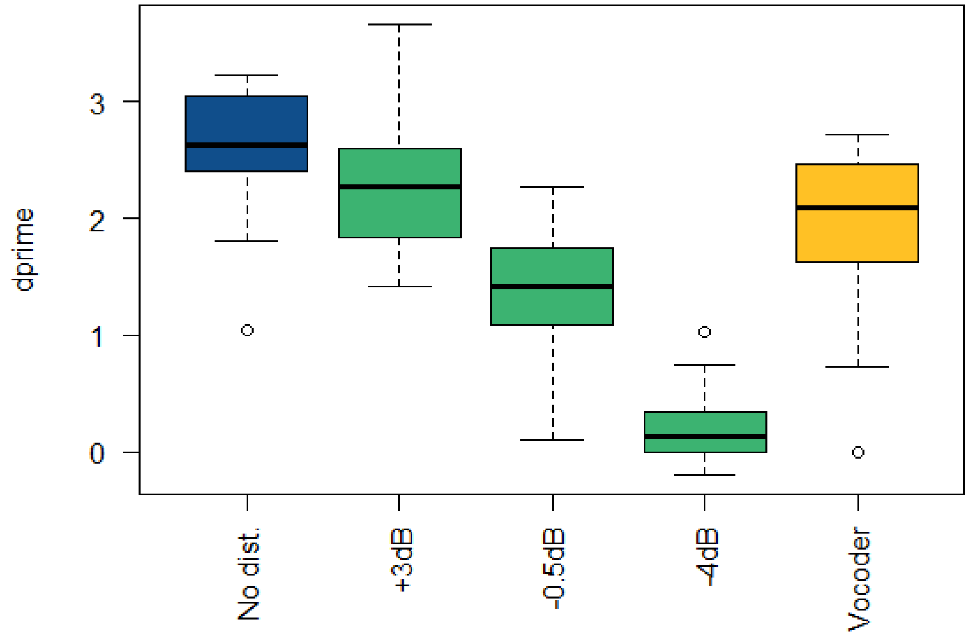
### 2.3.1 Statistical analysis

All electrophysiological results were analysed on three factors: d' as a measure of accuracy (intelligibility), whether there was degradation present in the target stimulus or not (normal vs. degraded), and what type of degradation (noise vs. vocoding). The latter two factors were contrast coded. To all neurophysiological data, linear mixed-effects models were applied, using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) in RStudio version 1.1 (R Development Core Team, 2013). Subject was used as a random factor for all analyses.

### 2.3.2 Behavioural data

Figure 2-1 shows d' (a bias-free measure of the ability to detect the difference between normal and anomalous sentences) in all five conditions. Intelligibility was highest for sentences without additional processing, and deteriorated with increasing noise. For the vocoded stimuli, intelligibility was approximately comparable to noise levels of +3dB. A linear mixed-effects model with condition as a main effect and subject as a random factor showed a statistically significant effect of condition ($\chi^2$ (4) = 160.27, p < 0.0001). A post hoc test with Bonferroni-Holm correction revealed that all the differences between conditions were statistically significant. These results indicate that the addition of noise as well as vocoding the target signal decreased intelligibility significantly.

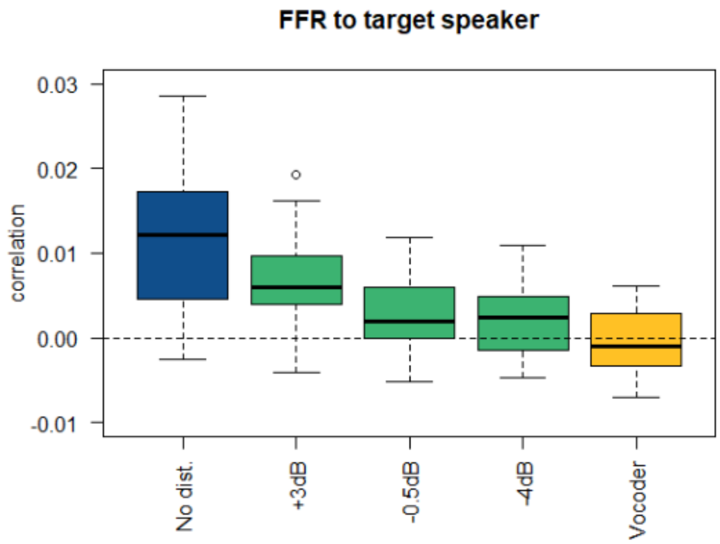
**Figure 2-1: Intelligibility (d-prime) scores for no processing on the target speaker ("No dist."), and 4 different levels of signal distortion (Noise: +3dB, -0.4dB and -4dB SNR, 14-channel vocoder)**




### 2.3.3 Frequency following response

Figure 2-2 shows boxplots of the FFR to the target and distractor speaker. FFR to the target speaker was relatively high for the non-degraded condition, but dropped away with increasing noise, similarly to cortical entrainment. For the vocoded condition, FFR approximated zero, which is to be expected: FFR is believed to track fundamental frequency (pitch), which is eliminated in a vocoded stimulus. Consequently, no FFR to the stimulus should be measurable in this condition. The FFR to the distractor speaker remained consistently high across all conditions, comparable to FFR tracking of the target speaker in the non-degraded condition. These results indicate that the addition of degradation to the signal strongly degraded FFR. At the same time, selective attention did not result in higher FFR tracking of the target speaker or suppression of the distractor. Thus, the primary effect observed here is the presence of degradation, and the significant difference in processing between noise and vocoding. Statistical analysis for the target speaker showed a significant effect of intelligibility (d') ($\chi^2$ (1) = 14.163, p < 0.0001), degradation ($\chi^2$ (4) = 25.798, p < 0.0001 and degradation type, $\chi^2$ (1) = 17.987, p < 0.0001). There was a significant interaction between d' and the presence of degradation ($\chi^2$ (1) = 25.798, p < 0.0001). There were no significant differences between conditions for FFR to the distractor speaker. These statistical results confirm the visual interpretation that degradation as well as the specific type of degradation significantly impacted FFR tracking activity, while selective attention did not.

**Figure 2-2: Signal reconstruction performance for FFR to target and distractor speaker (target speaker without processing and 4 levels of signal degradation, distractor speaker without any added degradation)**



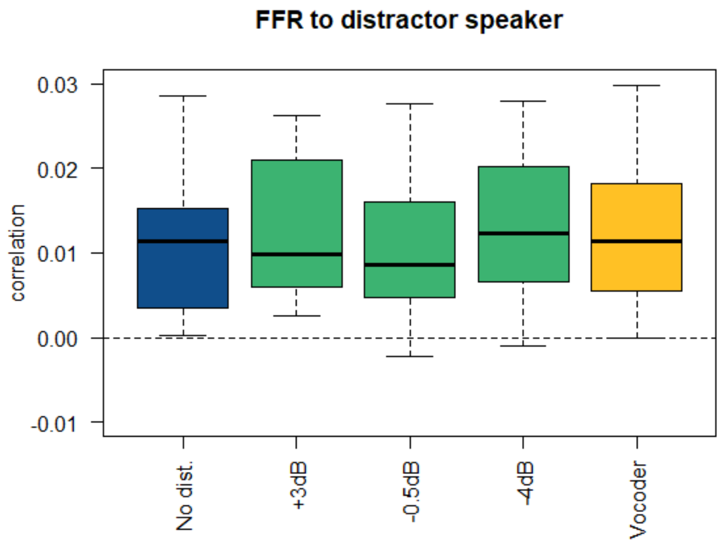
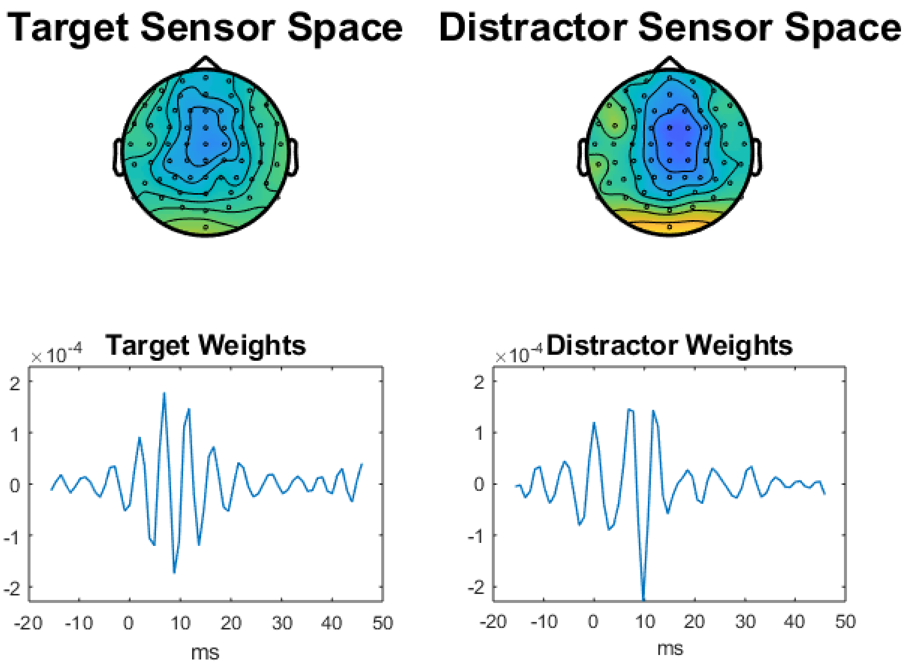



Figure 2-3 shows sensor space plots for target and distractor speakers as well as mTRF model weights. These were used for the stimulus-response mappings during the training phase and for the reconstruction of each stimulus (best electrode = Cz). The time lag between the trigger and audio was about 0.8ms, due to the amplification circuitry and the length of the insert earphone tube. After accounting for this lag, the model weights for signal reconstruction were strongest with a latency of 8-10ms. This means that these latencies had the greatest importance in terms of predictive power for the stimulus reconstruction, and indicates that the most meaningful response originated at these latencies. The sensor space plots suggests strong similarities between target and distractor speaker weights, with central channels showing the highest contribution to the model's parameters estimation. This seems to indicate that the processing of target and distractor were highly similar both in terms of latency and origin of generation.

**Figure 2-3: Target and distractor sensor space plots (at maximum value) and mTRF model weights for FFR reconstruction**




## 2.3.4 Cortical entrainment

Figure 2-4 displays coherence for target (in blue) and distractor speaker (in red). In Table 2-1, mean coherence values for target and distractor speakers are listed. For the normal (i.e. non-degraded) condition, there was a clear target speaker entrainment advantage in the delta and theta (2-8Hz) range, as previously has been recorded in other two-talker entrainment studies. This advantage was still clearly visible for the +3dB SNR and -0.5dB SNR condition, but dropped away at -4dB SNR. There was also no additional entrainment to the target speaker in the vocoder condition. For statistical analysis, mean values in the frequency region of 2-8 Hz were extracted for target and distractor for each participant and each condition. The mean distractor value was subtracted from the mean target value to compute the difference. The analysis revealed that there was a significant effect of intelligibility, ($\chi^2$ (1) = 9.9245, $p < 0.001$), as well as degradation type ($\chi^2$ (1) = 18.978, $p < 0.0001$). Presence of degradation alone did not affect the measure significantly ($\chi^2$ (1) = 0.2286, $p = 0.6326$). There were no significant interactions between the factors. These results indicate that, while target speaker entrainment remained relatively robust to noise, high noise levels led to a drop in entrainment in parallel with low intelligibility levels. For the vocoded condition, however, entrainment to the target speaker was also strongly reduced despite high levels of intelligibility. This means that, in contrast to the noise conditions, vocoding resulted in a dissociation between intelligibility and cortical entrainment. More generally, this surprising result is an indication that, while intelligibility tends to co-occur with cortical entrainment in certain conditions, the relationship is not closely coupled for all types of degradation.

**Figure 2-4: Signal reconstruction performance for cortical entrainment to target and distractor across all conditions, without target signal processing and 4 levels of degradation (top); boxplots and beeswarm plots to show variability of the data (bottom)**

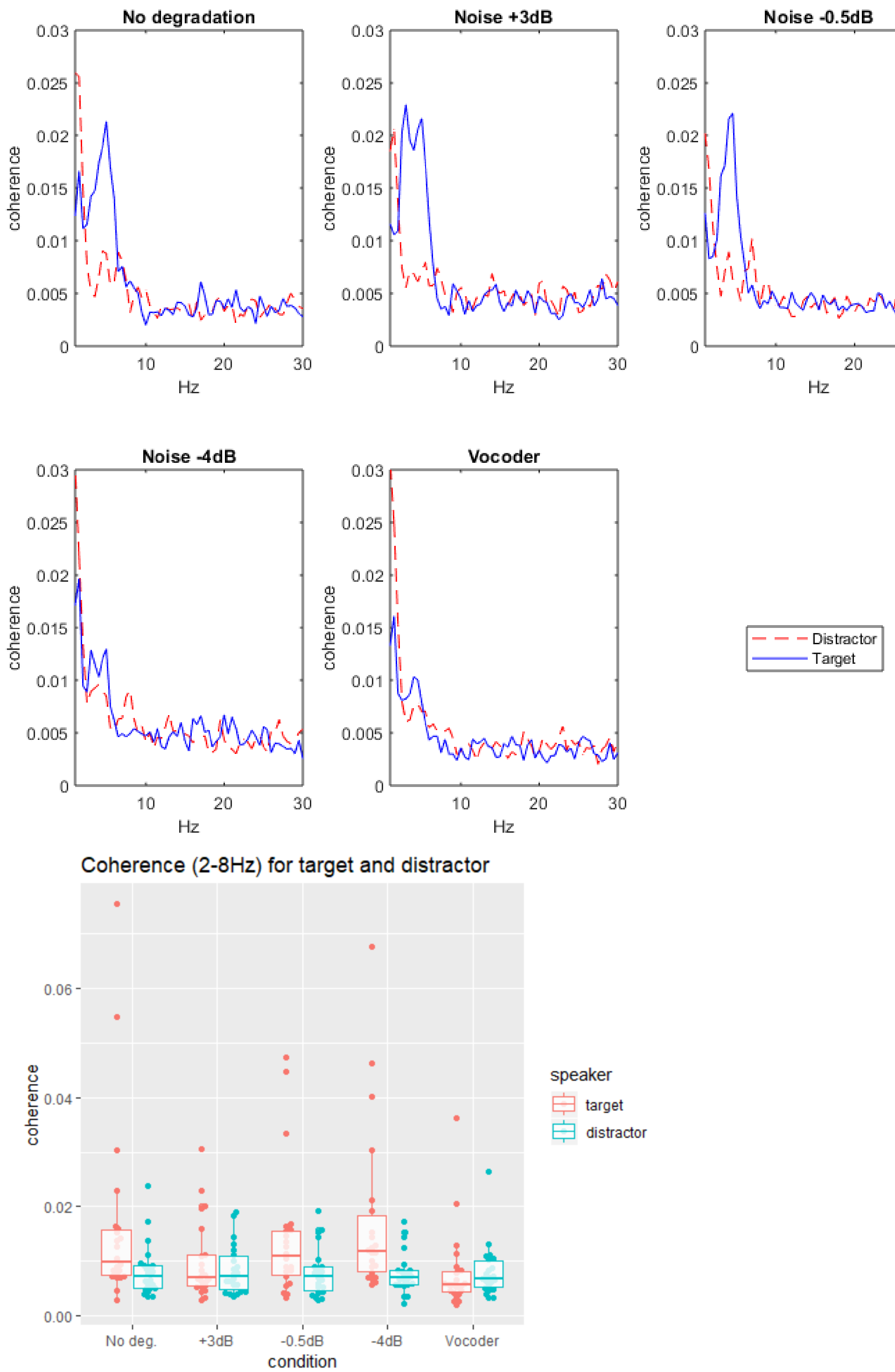


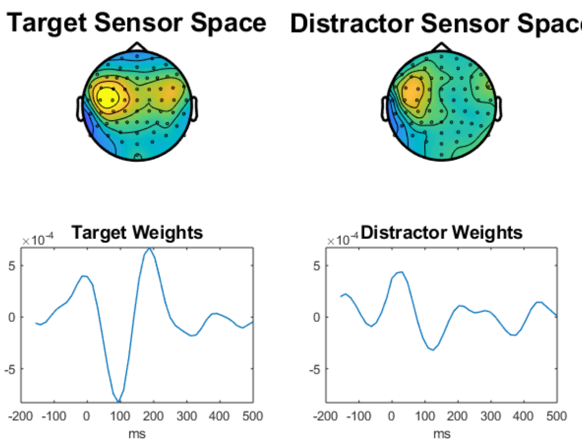
**Table 2-1: Mean coherence values to the target and distractor speaker per condition**

| **Mean coherence value (target) 2-8Hz** | |
|---|---|
| No degradation | 0.0164 |
| Noise, +3dB SNR | 0.0174 |
| Noise, -0.5dB SNR | 0.0144 |
| Noise, -4dB SNR | 0.0100 |
| Vocoder | 0.0079 |

| **Mean coherence value (distractor) 2-8Hz** | |
|---|---|
| No degradation | 0.0083 |
| Noise, +3dB SNR | 0.0079 |
| Noise, -0.5dB SNR | 0.0080 |
| Noise, -4dB SNR | 0.0085 |
| Vocoder | 0.0081 |

In Figure 2-5, sensor space plots and mTRF weights are displayed for both target and distractor speaker (best electrode = FC3). The sensor space plots seemed to reflect a left-lateralized component, both for the target and the distractor speaker model. The left-lateralisation might be due to the experimental presentation (dichotic listening task with target speaker in the right ear), and might suggest enhanced processing of the target and it being reflected in the characteristics of the latency and localisation of the model's stimulus-response mapping. The latencies of the target model seemed to resemble the N1-P2 component measured in typical ERP experiments.

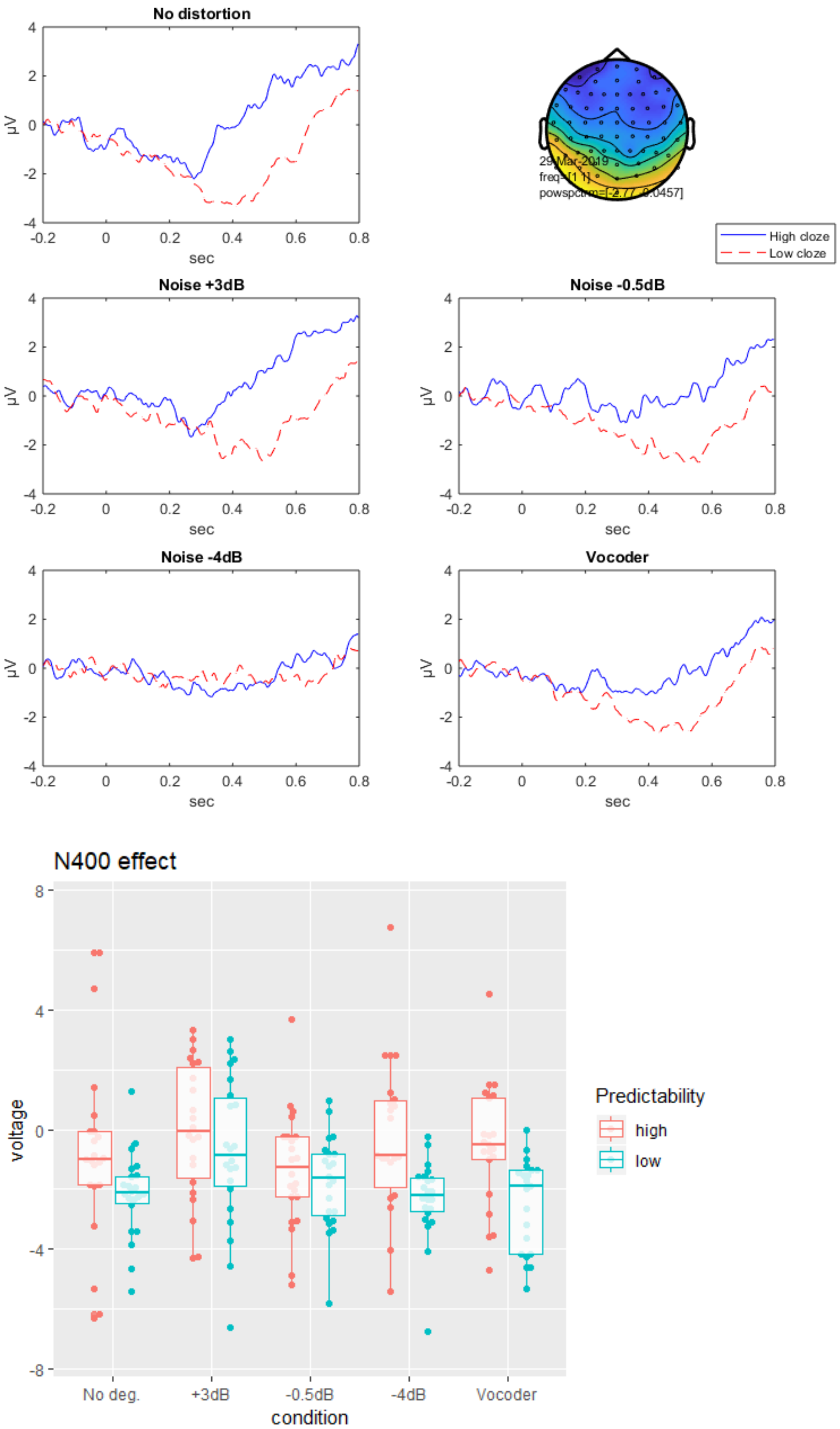
**Figure 2-5: Target and distractor sensor space plots (at maximum value) and mTRF model weights for cortical entrainment**



### 2.3.5 N400

Figure 2-6 shows voltage over time for the five conditions, with the high cloze probability sentences in blue and low cloze probability sentences in red. In the non-degraded condition, a clear N400 effect was observed (i.e. voltage dropped more around 400ms for low cloze probability sentences than it did for high cloze probability sentences). With increasing noise, the N400 effect became smaller, and disappeared in the highest noise condition (-4dB SNR). In the vocoded condition, the N400 effect observed was comparable to the easy or medium noise condition. For statistical analysis, average values for each participant in each condition during a time window of 300ms to 500ms were extracted. The analysis revealed a significant effect of d' ($\chi^2$ (1) = 7.98, $p < 0.01$), but only for low cloze probability sentences. There was no significant effect of presence of degradation ($\chi^2$ (1) = 0.893, p = 0.345), nor for type of degradation ($\chi^2$ (1) = 0.993, p = 0.3190). These findings overall relate well to behavioural results and indicates that the N400 is related to intelligibility in all conditions. They support the visual interpretation that significant differences between conditions can be attributed largely to differences in intelligibility, regardless of the type of degradation. These results thus stand in contrast to those from cortical entrainment, which showed differences between degradation types (noise vs. vocoding) despite similar levels of intelligibility.

**Figure 2-6: N400 effect for target speaker sentences with high and low cloze predictability, without added processing and 4 levels of signal degradation (+3dB, -0.5dB, -4dB SNR) (top); boxplots and beeswarm plots to show variability of the data (bottom)**



## 2.4 Discussion

This study investigated the effect of different types and degrees of masking on the intelligibility and form of the target speech. Peripheral and central measures of neural speech processing were impacted differently by various types and degrees of degradation. Cortical entrainment also revealed different processing of speech presented in noise vs. vocoded speech, despite similar levels of intelligibility.

A central finding was the varying influence of different factors on speech processing from peripheral through to central processing units. The earliest stage of neural processing measured (FFR) was highly sensitive to the degree to which the stimulus was intelligible, as well as presence and type of degradation. The next measure along the neural pathway, cortical entrainment, was significantly influenced by type of degradation and intelligibility, but not the mere presence of degradation alone. N400, as a late processing stage, was associated with intelligibility, but not presence of degradation or degradation type. These differences between the measures suggest that peripheral processing is more strongly driven by attributes of the stimulus itself, whereas central semantic processing is more closely related to higher-order factors such as intelligibility. A possible explanation might be that at the brainstem level, the first basic distinction to be made is the separation of the incoming signal into various streams. Consequently, the FFR is highly sensitive to the presence of any type of signal degradation. At the cortical level, then, these streams are analysed in more detail with regards to the type of degradation. Finally, at the lexical as a central processing level, high-level semantic information that is dependent on intelligibility, such as word context and predictability (Kutas & Federmeier, 2000), drive processing.

Previous research has extensively investigated the extent to which unattended stimuli are processed automatically in the brain (Lavie et al., 2004), and at what point the listener's intention to attend to a specific stream in the presence of informational masking ('top-down' attention) actively shapes auditory processing. While a growing body of evidence suggests that at the cortical level, an active neural mechanism enhances target sounds and suppresses distractor sounds (e.g. Ding & Simon, 2012, 2013; Kerlin et al., 2010), the effects of selective attention on the (much earlier) brainstem response are less clear. While one previous study (Forte et al., 2017) has, already at this subcortical processing stage, found a target speaker entrainment advantage similar to that observed in cortical entrainment studies, others have presented evidence that the FFR as an early measure might be pre-attentive and thus follows voice pitch invariably even when the listener does not actively attend to a sound stream (Holmes, Purcell, Carlyon, Gockel, & Johnsrude, 2018; Varghese, Bharadwaj, & Shinn-Cunningham, 2015). In the present study, the cortical measure showed a clear entrainment advantage for the attended speaker in low frequency regions, similar to findings from previous studies. In contrast, no such difference was observed between attended and unattended speaker at the brainstem level. While tracking of the unattended speaker remained high throughout all conditions, FFR to the target speaker was significantly affected by noise degradation and absent in the vocoded condition. Although these findings might thus suggest that the brainstem response is pre-attentive, an important confound in the study was that all participants invariantly only attended to the female speaker, who naturally had a higher fundamental frequency than the male distractor speaker. Since it is known that the brainstem response decreases in magnitude with higher frequencies (Tabachnick & Toscano, 2018), it remains unclear whether the effect observed is masked by a lower baseline response to the higher f0 of the female talker. Another complicating factor in the present study is that the degradation (noise and vocoding) was only added to the target speaker channel, but not the distractor. Possibly, equalizing the conditions for both speakers (i.e. introducing noise and vocoding for the distractor as well) might have allowed for the observation of attention effects at this subcortical processing stage already.

Whereas results from this study indicated that the FFR was highly sensitive to degradation by noise, cortical entrainment to the target remained relatively robust until a high level of noise, but then fell off (SNR -4dB). While the robustness of cortical entrainment to noise degradation has been demonstrated before (Ding et al., 2014), studies on the FFR have equally concluded that, compared to other portions of the brainstem response, it remains relatively robust to noise (Cunningham et al., 2001; Russo et al., 2004). However, previous studies have measured the FFR mostly in positive – and at most neutral – SNRs. One such study (Russo et al., 2004) showed that the FFR amplitude was reduced by about 30-40% at SNR levels of +5dB. In the present study, correlation was found to be reduced by 42% at an SNR of +3dB (and reductions of 78% and 81% at SNRs of -0.5 and -4dB, respectively). Considering that the easiest noise level used here was slightly worse than +5dB, the reduction in FFR as a response to noise is thus highly similar to previous findings in the literature. While this reduction might still be considered robust relative to onset portions of the brainstem response, in comparison with cortical entrainment the FFR appears much less resistant to degradation in noise.

Unsurprisingly, no FFR to the vocoded stimulus type was measured; this was to be expected, since fundamental frequency is eliminated in vocoded material (which is precisely the aspect of speech FFR is thought to track, Qin & Oxenham, 2005). Interestingly, however, there was also no target talker cortical entrainment advantage observed for the vocoded condition, even though intelligibility levels were relatively high. In an MEG study with noise and vocoders, Ding & Simon (2014) showed that while vocoding in quiet did not significantly impact neural entrainment, adding noise (+3dB SNR) to vocoded stimuli severely disrupted entrainment. It has also been reported (Friesen et al., 2001) that reduced spectral resolution affects speech intelligibility severely in noisy conditions only, but not in quiet. In the present study, there were no vocoded stimuli in quiet, as the distractor speaker was always presented simultaneously in the other ear. These results suggest that neural entrainment to vocoded speech is highly sensitive to degradation, even if the 'noise' is a single distractor speaker and intelligibility is relatively high with a 14-channel vocoder. The comparison with results from the noise conditions point to potentially different processing mechanisms for these two types of stimulus degradation, which might be due to their respective characteristics: while noise is added to the stimulus ('source-external' degradation), for vocoding the stimulus itself is altered ('source-internal' degradation). This difference might consequently influence the way in which the brain attempts to filter out relevant information – while noise as an energetic masker might be more easily disassociated from the target stimulus and represented as a separate auditory object (Shinn-Cunningham, 2008), vocoded material might be subject to a more complex processing mechanism due to the inseparability of the degradation from the stimulus.

Compared to cortical entrainment and FFR, the N400 effect was more closely related to intelligibility than stimulus characteristics such as level or type of degradation. As the N400 occurs relatively late, it might as such be less driven by bottom-up, 'automatic' processes such as sounds irrelevant to the listener's intentions (e.g. a distractor talker or background noise; Golestani, Hervais-Adelman, Obleser, & Scott, 2013).With increasing noise levels, the N400 effect became smaller and disappeared in the highest noise condition (-4dB SNR). As previous research has indicated that a greater N400 can be associated with more effortful semantic integration and higher listening effort (Chwilla et al., 1995; Kuipers & Thierry, 2011; Song & Iverson, 2018), the results can be interpreted as follows: as listening difficulty increased due to higher noise levels, more lexical search activity was necessary for otherwise highly predictable words, thus enlarging the N400 response. At the same time, less lexical search activity might have been performed for words that are not highly predictable from the semantic context, as the brain is less able to process the sentence and at some point stops searching for matches in the lexicon. This might have reduced the N400 response to low predictability sentences. The combined effect of an enlarged N400 for high predictability sentences and a decreased N400 for low predictability sentences could consequently have caused the progressive reduction of the N400 effect in increasingly adverse conditions.

In sum, this study showed that various types and degrees of degradation have different effects on neural measures of speech processing. While peripheral processing was most strongly affected by bottom-up factors such as stimulus characteristics, central lexical processing was most closely associated with intelligibility. It was also demonstrated that noise and vocoding as different types of degradation might be processed in different ways; whereas cortical entrainment to the target speaker remained high at good intelligibility levels, no such additional entrainment to the target was measured in the vocoded condition, despite similar levels of intelligibility.

# Chapter 3: Neurophysiological measures of attention in hearing aid users

## 3.1 Introduction

Various signal processing techniques are commonly employed in hearing aids in order to enhance listener experience and speech recognition performance in difficult listening situations. Among these, two of the most prominent signal processing algorithms implemented in hearing aids are noise reduction and directional microphones (beam formers) (Bentler, 2005). Noise reduction algorithms analyse the acoustic environment and seek to reduce the gain in frequencies that are dominantly populated by noise rather than the signal. Directional microphones are used to amplify sounds from a specific direction, usually from the front, whilst reducing sounds from other directions.

One particular problem with the performance measurement of technologies routinely implemented in hearing aids is that it is difficult to capture the value to the user in traditional measures such as speech recognition performance. While some studies with noise cancelling techniques and directional microphones report better speech recognition performance when these programs are switched on (e.g. Ricketts & Hornsby, 2006; Walden, Surr, Cord, Edwards, & Olson, 2000), these results have been far from unequivocal. Other studies have found that users still fail to regularly switch to the optimal program for a specific acoustic environment (Ricketts et al., 2017), or that there is no actual performance improvement in speech recognition, but rather solely a subjective user preference towards these programs (Desjardins & Doherty, 2014; Ricketts & Hornsby, 2005). Therefore, it is necessary to explore other ways of measuring speech perception in addition to the traditional measures of behavioural performance and the pure tone audiogram. Neural measures of speech processing could enhance our understanding of how phenomena such as attentional focus and listening effort are affected by hearing loss, and could allow us to capture the value of hearing aid features in dimensions beyond pure speech intelligibility.

In the present study, several of the measures from the toolbox developed in Study 1 were applied in order to investigate neural speech processing and auditory attention in hearing-impaired people. On top of the neurophysiological measures, subjective ratings of listening effort and a behavioural measure of speech comprehension were collected. A normal-hearing and a hearing-impaired group were tested on a speech-in-noise task, and EEG activity was recorded during the experiment. Subjects were fitted with a hearing aid (HI subjects for hearing loss compensation, and NH subjects with minimal amplification for better comparison) and three different programmes: hearing loss compensation only, noise reduction, and directional microphone (beam forming).

### 3.1.1 Speech perception and selective attention in older and hearing-impaired adults

Hearing-impaired people not only suffer from a poorer representation of the incoming auditory signal due to cochlear hair cell damage, they are also less able to robustly direct attention to a specific sound stream while ignoring distractors. A study with children with and without hearing loss (aged 7 to 16 years) found that the hearing-impaired group had less EEG activity indicative of selective attention compared to their normal-hearing peers (Holmes, Kitterick, & Summerfield, 2017). This finding suggests that already at a young age, hearing impairment affects neural processing and auditory attention adversely. Another study has also shown that the hearing-impaired performed significantly poorer on a behavioural measure of selective auditory attention, and had less robust EEG representations, even when their age was not significantly different from the normal-hearing group (Dai et al., 2018). Together, these results point to underlying differences in attentional selection as a result of hearing impairment. However, it remains unclear whether these differences in attentional selection are due to impaired cognitive function ('top-down' factors), or a consequence of normal cognitive processes operating on a distorted representation of the incoming acoustic signal ('bottom-up' factors).

Age – even in the absence of a hearing impairment measurable through pure tone audiometry – is related to changes in cognitive processes that may interfere with speech processing, phase-locking and selective auditory attention at various levels (Shinn-Cunningham & Best, 2015). For example, age has been linked to declines in cognitive factors, specifically the encoding of temporal events (Gordon-Salant, 2006). Based on these findings, studies have shown that this less precise temporal coding results in deficits related to temporal fine structure processing in older adults, despite normal thresholds of hearing (Fuellgrabe, 2013; Grose & Mamo, 2010). EEG research indicates that neural speech tracking is also affected by age, with several studies showing increased cortical envelope encoding compared to younger subjects, at least in certain frequency regions (Goossens, Vercammen, Wouters, & Wieringen, 2016; Presacco, Simon, & Anderson, 2016a, 2016b). This greater encoding of the speech envelope might result from a similar unbalanced processing mechanism of temporal envelope and fine structure cues as that observed in hearing-impaired people. Overall, these findings suggest substantial changes in auditory processing with age, even despite the absence of hearing impairment.

Behaviourally, older adults (62.2 years average age, compared to 25.2 years average age for younger group) also performed worse on a divided attention task (Getzmann, Golob, & Wascher, 2016). There is further indication that ageing decreases the ability to suppress the lexical activation of unwanted distractors: for example, older adults are more adversely affected than young adults by a distractor that is meaningful speech than randomly ordered speech or a foreign language (Tun, O'Kane, & Wingfield, 2002); and they display higher error rates and longer reaction times overall compared to the younger group, due to the inability to suppress distractor speech and its likely interference at a lexical level (Oberem, Koch, & Fels, 2017). An earlier fMRI study suggests that two of the reasons for older adults' decreased ability to comprehend speech in adverse conditions could be (1) reduced activation in certain brain areas such as those associated with language-processing, and (2) a decrease in coordination between regions involved in speech comprehension (Peelle, Troiani, Wingfield, & Grossman, 2010). All these findings indicate that even in normal-hearing older adults, speech comprehension is adversely affected due to changes in central systems that become less efficient at selective attention. Taken together, hearing loss and ageing potentially result in the compounding effect of poorer speech representation at peripheral levels as well as less effective central processing responsible for separating speech streams, directing attention to a target, and suppressing distractors.

### 3.1.2 Cortical entrainment and its relationship with age and hearing impairment

As ageing has been associated with a variety of changes in the brain and cognition (Peng et al., 2018), it is worth investigating its effects on cortical entrainment. Previous MEG and EEG studies with auditory evoked potentials have found that older adults exhibit larger amplitudes (Alain, McDonald, & Van Roon, 2012; Sörös, Teismann, Manemann, & Lütkenhöner, 2009) and latencies (Tremblay, Piskosz, & Souza, 2003) in response to syllables and complex tones, compared to a younger group. Yet, a recent study on non-speech sounds found that neural entrainment is weakened in older adults, and that the attention effect (as the difference between active and passive listening) is larger in younger than older adults, consistent with the behavioural observation that selective auditory attention seems to decline with age (Henry, Herrmann, Kunke, & Obleser, 2017). Results from the few studies using continuous speech (as opposed to non-speech sounds, syllables or tones) and the stimulus reconstruction technique applied in this thesis, were consistent with previous EEG findings measuring auditory evoked potentials: these recent entrainment studies have shown that older adults display exaggerated entrainment compared to younger adults (Presacco et al., 2016b, 2016a). They also found that higher reconstruction accuracy was associated with worse behaviour in a visual flanker task (Presacco et al., 2016a), a measure of inhibitory control and attention (Weintraub et al., 2013). This task specifically measures a subject's ability to inhibit visual attention to irrelevant stimuli. Due to these results, the authors of this study hypothesized that the increased difficulty of understanding speech in adverse conditions for older adults may at least partly be due to a reduced ability to supress an unattended stimulus. In summary, it appears that while in young, normal-hearing adults higher entrainment is associated with positive effects (more attention, higher intelligibility), this relationship seems to be reversed in older, hearing-impaired adults.

Findings from the effects of hearing loss on neural entrainment point into a similar direction as for ageing. Animal studies have determined that hearing loss increases the temporal precision and can amplify the representation of the envelope by up to 50% (Henry, Kale, & Heinz, 2014; Kale & Heinz, 2010, 2012). The few studies that have been conducted so far on neural entrainment and hearing impairment on the whole suggest a similar mechanism in humans. An MEG study on older people (around 60 years old) found hearing loss to be associated with higher neural speech tracking and worse speech perception in noise (Millman, Mattys, Gouws, & Prendergast, 2017). Another EEG study, using tones for acoustic stimulation and measuring auditory steady-state responses (evoked auditory potentials), found that while in young (20-30 years old) and middle-aged (50-60 years old) adults, hearing impairment was associated with enhanced neural synchronization to the stimulus (relative to their normal-hearing controls), older adults (70-80 years old) showed similar patterns in neural representation, regardless of hearing loss (Goossens, Vercammen, Wouters, & van Wieringen, 2018). These effects of enhanced, rather than reduced, tracking of the temporal envelope observed in both ageing and hearing-impaired people might be due to the reduced processing of temporal fine structure cues, and therefore greater reliance on envelope cues (Moore et al., 1996; Sek et al., 2015). Another EEG study on hearing-impaired people (Petersen, Wöstmann, Obleser, & Lunner, 2017) also found that while older adults show the same attentional entrainment advantage for attended speech as younger adults, worse hearing corresponds to a reduction in that advantage. In other words, the difference in neural entrainment between an attended and an ignored talker becomes smaller with increasing hearing loss. Notably, this decrease in tracking difference between target and distractor is not due to the reduction in entrainment to the target speaker, but rather the increase in tracking of the distractor speaker. The authors therefore suggest that hearing impairment might be associated with a decrease in the inhibition of the irrelevant stimulus – a similar effect as found with age (Oberem et al., 2017). In sum, age and hearing impairment might lead to processing imbalances of temporal speech cues, as well as the less effective suppression of distractor cues – both of which have been shown to influence cortical entrainment.

### 3.1.3 Intersubject correlation and auditory attention

Apart from measuring the similarity between the auditory and the neural signal (i.e. neural entrainment), it is also possible to measure the similarity of the neural signal across subjects while performing a task. It has been shown in fMRI studies (Hasson, Furman, Clark, Dudai, & Davachi, 2008; Hasson, Malach, & Heeger, 2010; Wilson, Molnar-Szakacs, & Iacoboni, 2008) that hemodynamic responses to audiovisual stimuli can be reliably reproduced across viewers. To investigate the intersubjective reliability of neural responses to auditory stimuli with a more precise temporal resolution, recent studies have conducted MEG (Lankinen, Saari, Hari, & Koskinen, 2014) as well as EEG studies (Cohen & Parra, 2016; Dikker et al., 2017; Dmochowski et al., 2014; Dmochowski, Sajda, Dias, & Parra, 2012; Ki, Kelly, & Parra, 2016; Poulsen, Kamronn, Dmochowski, Parra, & Hansen, 2017). Intersubject correlation

(ISC) is computed similarly to traditional evoked potentials in that both measures have higher magnitude for more reliably reproduced stimuli (for evoked potentials across trials, for ISC across subjects). For studies involving memory and attention, however, ISC has the advantage that the same stimulus does not need to be presented to the same subject repeatedly, as averaging is done across, but not within, subjects (Cohen & Parra, 2016). More specifically, ISC can be evaluated through correlated component analysis (Dmochowski et al., 2014, 2012). Correlated component analysis is similar to principal component analysis, and is designed to find linear combinations of electrodes consistent across subjects and maximally correlated with one another. The main difference between correlated component analysis and principal component analysis is that correlated component analysis extracts data components with maximal correlation (rather than maximal variance) (Cohen & Parra, 2016).

Higher ISC has been associated with a variety of factors such as superior memory encoding (Cohen & Parra, 2016), enhanced viewer engagement (Dmochowski et al., 2012), high teaching impact (Poulsen et al., 2017) and successful communication (Stephens, Silbert, & Hasson, 2010). One recent study (Ki et al., 2016) explicitly manipulated the attentional state of listeners and compared ISC between an active and a passive listening task. ISC was significantly enhanced when subjects actively attended to the stimulus. The authors also computed alpha power, a more conventional measure of auditory attention, and found that ISC provided more predictive power as a measure of attention, while at the same time being less susceptible to individual differences.

Studies with audiovisual stimuli have demonstrated that neural variability changes with age. While one such study concluded that adults have more similar neural responses to audiovisual stimulation than children (Cantlon & Li, 2013), another study with adults from age ranges of 18 to 88 found that ISC declines with age (Campbell et al., 2015). The most recent study provided further supportive evidence that ISC declines with age, and confirmed previous findings that ISC is higher for more engaging stimuli (Petroni et al., 2018). No studies so far have been conducted to investigate how hearing impairment impacts ISC. If ISC is indeed a measure of auditory attention, and hearing impairment coincides with a reduced ability to direct and sustain selective attention to a target stream, it may be assumed that ISC declines with hearing loss. However, due to the mixed evidence on the influence of age on ISC, this relationship is only tentative and needs to be corroborated by further neural evidence.

The present study investigated differences in neural language processing between a young, normal-hearing and an older, hearing-impaired population, and the effects of signal-enhancing techniques commonly employed in hearing aids. Both groups were tested on a listening task with different levels of noise and two different settings on the hearing aids. Two neurophysiological measures of attention were extracted from the data: cortical entrainment and intersubject correlation. Behavioural responses as well as subjective ratings of listening effort were collected. With this study, the following questions were examined: How do neurophysiological measures of attention differ between the NH and HI group? Do noise reduction and directional microphone mode influence subjective, behavioural and neurophysiological measures, and if so, how?

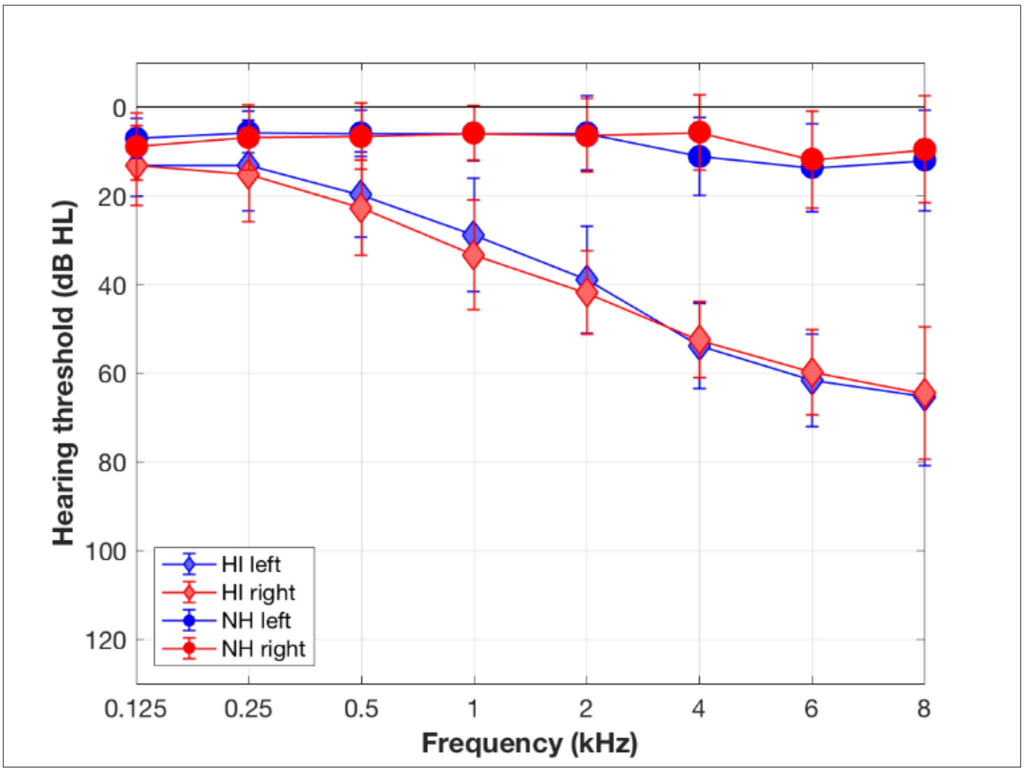
How are different measures affected differently in the two groups, and between different conditions, and how do they relate to one another?

## 3.2 Materials and Methods

### 3.2.1 Subjects

Twenty-two hearing-impaired adults aged between 47 and 78 years old were tested (8 = female, 14 = male, mean age = 67.5). All of these participants suffered from mild to moderate bilateral sensorineural hearing loss, defined as follows: 1) air conduction thresholds <60 dB hearing level from 125 to 4,000 Hz bilaterally and 2) no interaural asymmetry (>15 dB hearing-level difference at no more than 2 adjacent frequencies). Subjects both with and without prior HA experience participated in the study. For the control group, 22 normal-hearing adults aged between 31 and 50 years old were tested (5 = female, 17 = male, mean age = 37.9). These participants had normal-hearing, defined as follows: 1) air conduction thresholds ≤25 dB hearing level from 125 to 4,000 Hz bilaterally and 2) no interaural asymmetry (15 dB hearing-level difference at no more than 2 adjacent frequencies) (British Society of Audiology, 2017).

**Figure 3-1: Hearing thresholds across 0.125-8kHz for the normal-hearing and the hearing-impaired groups. Blue = Left, Red = Right**




All participants were native Swiss German speakers, recruited from the Zurich area, Switzerland. The study was approved by the local ethics committee (Kantonale Ethikkommission Zürich). The subjects provided their informed consent and were compensated for their time and travel expenses.

### 3.2.2 Hearing aid

Participants were bilaterally fitted with Phonak Audéo B90-Direct receiver-in-canal (RIC) hearing aids coupled to disposable canal earmolds with no venting (power domes). These were used to provide maximum occlusion and increase the efficacy of the programs tested in this study (Magnusson, Claesson, Persson, & Tengstrand, 2013). For the hearing-impaired group, hearing aid gain was determined on the bases of participants' individual audiograms and calculated automatically by the manufacturer's fitting software (Phonak Target software). Three different settings on the hearing aids were used in this experiment: (1) amplification to compensate for the subject's individual hearing loss, imitating the spatio-acoustic characteristics of a human ear ('real ear setting'), (2) noise reduction algorithm (noise reduction, NR), and (3) directional microphone mode (DM). For the normal-hearing group, the same settings were used. However, amplification was fixed to a minimum level (20dB 'hearing loss') to ensure the algorithms' (noise reduction and directional microphone) activation in the respective conditions.

### 3.2.3 Stimuli

Materials were extracted from a radio show presented in the Swiss National radio program SRF (Schweizer Radio und Fernsehen, SRF 2018). The show ("Zwischen den Schlagzeilen", "Between the headlines") presents in-depth interviews with Swiss journalists reporting from various regions all over the world. Each segment was taken from a different episode and had a duration of approximately 4 to 4.30min. In total, participants heard 12 different segments ('target stories'). Two Swiss radio presenters discussed a current topic from a specific country or part of the world. Stimuli were spoken in Standard German with a Swiss accent. Episodes were selected to avoid overly political or potentially upsetting content. The experimental set up consisted of 12 loudspeakers placed in a circle around the subject. The target stimulus was always presented at 65dB SPL from the loudspeaker facing the subject from the front. For each story, 11 distractor files (multi-talker unintelligible babble noise) were created to be played from all loudspeakers except the one playing the target stimulus. These distractor files consisted of a mixture of all stories except for the one currently presented to the subject as the intelligible target, and created an unintelligible babble noise to mask the target. Two different signal-to-noise ratios between target and maskers were used: 0dB SNR, and +5dB SNR. All target and distractor files were matched in duration, so that sound onset coincided. In total, participants were presented with six different conditions:

a) +5dB SNR (hearing loss compensation only)
b) +5dB SNR, NR (noise reduction)
c) +5dB SNR, DM (directional microphone)
d) 0dB SNR
e) 0dB SNR, NR
f) 0dB SNR, DM

Each participant was assigned to one of two groups. For each group, the six conditions were randomly assigned to certain stories. Each participant heard two stories in each condition.

### 3.2.4 Apparatus

EEG data were collected at a sampling frequency of 2,048 Hz using the BioSemi Active Two system (BioSemi, Amsterdam, Netherlands) with 64 (Ag/AgCl) electrodes. 7 reference electrodes were placed on the face to control for eye blinks and movement as well as muscle-induced artefacts. Two electrodes placed on the mastoids were used as reference. Participants were seated in a sound-attenuating booth. The experiment took approximately one hour to complete.

### 3.2.5 Procedure

Participants first performed a standard tone audiometry test in order to determine their hearing thresholds for the right and left ear separately. After a visual inspection of the ear canal, hearing aids were fitted bilaterally. Power domes were used to allow for maximal occlusion of the ear canal. Amplification gain was set to 100%. A feedback test was performed to check the hearing aid's fit in the ear canal and to avoid unwanted microphone feedback during the experiment. Participants were then seated inside the booth on a chair surrounded by the 12 loudspeakers. They were instructed to only pay attention to the target conversation played from the front loudspeaker and disregard the noise from the other loudspeakers. After each story, participants were presented with three questions on a touchscreen placed next to them. The first two questions served as an indicator of intelligibility levels as well as a way to make sure participants did indeed pay attention to the target stories. These were comprehension questions with four answer options each, out of which only one was correct. The first question targeted general text comprehension and asked about the most appropriate title for the stories participants had just heard. The second question was more detailed and asked about a fact discussed towards the end of the story. After each of these comprehension questions, participants were given feedback on whether they had chosen the right answer. The third question was a question asking for subjective listening effort expended on this particular story. Participants could choose from 14 buttons ranging from “no effort” (‘mühelos’) to “noise only” (‘nur Störgeräusch’). The first story was always presented with a high SNR (+5dB) and real ear setting, so that participants could familiarize themselves with the task. All other stories were presented in randomized order. After each story, participants were given a short break.

### 3.2.6 EEG analyses

An EEG experiment was conducted (Biosemi Active Two system with 64 (Ag/AgCl) electrodes) with 7 reference electrodes placed on the face to control for eye blinks and movement as well as muscle activation-induced artefacts. Participants were seated in a sound-attenuated booth. The experimental task took approximately one hour to complete. Data were recorded at 2,048 Hz. Two electrodes placed on the mastoids were used as reference. Noisy channels were interpolated. For cortical analyses, data were filtered between 0.1 and 30 Hz and downsampled to 256 Hz. Filters used were band-pass zero-phase causal Butterworth filters of the 3$^{rd}$ order (response slope of -18dB per octave) as implemented in the ERPlab toolbox (Lopez-Calderon & Luck, 2014) of EEGlab (Delorme & Makeig, 2004). Independent Component Analysis was applied to correct for eye blinks and horizontal eye movements. All pre-processing procedures were performed in Matlab using the Fieldtrip toolbox (Oostenveld et al., 2011).

**Cortical entrainment.** As in Study 1, the mTRF Toolbox (Crosse et al., 2016) was used to generate multivariate Temporal Response Functions in order to train a backward model on the data. The EEG data were first matched with the respective amplitude envelope of each story. Then, the regression model was trained to analyse stimulus and response data for time lags ranging from 0ms to 400ms. The training set always consisted of all stories except one. The model performance was then tested on the remaining story (test set), reconstructing the amplitude envelope based on the EEG data from this story. These predictions from the model were then compared with the actual stimuli (i.e. amplitude envelopes of the stories) and coherence computed as a measure of cortical entrainment. This was done by segmenting the data in to 1-sec Hann windows with 50% overlap, and calculating coherence from the cross-spectral density of the FFT of the two signals, divided by the power spectrum of each signal.

**Intersubject correlation.** Intersubject correlation (ISC) was computed based on the toolkit provided by Cohen & Parra (2016). For ISC analysis, EEG data were downsampled to 64Hz. All data for each individual subject were concatenated in the same order, and data compiled across all subjects. Then, the cross-covariance between all subjects as well as within- and between-subject covariances were computed. Correlated components were calculated and sorted according to intersubject correlation power. Finally, ISC for the first component was computed for each condition separately.

## 3.3 Results

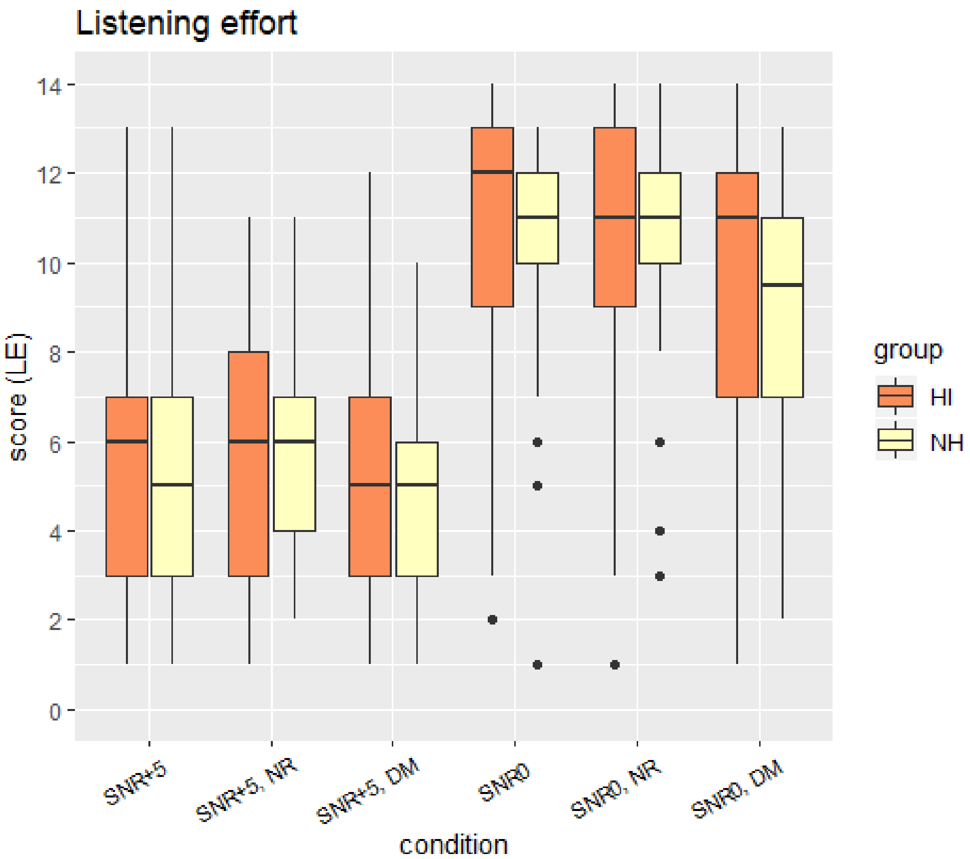
### 3.3.1 Statistical analysis

All statistical analyses were conducted in RStudio version 1.1 (R Development Core Team, 2013). Linear mixed-effects models were applied, using the lme4 package (Bates et al., 2014) and model comparisons to evaluate the significance of each factor as well as interaction effects. SNR level, type of algorithm and group were considered as main factors, and subject was treated as a random factor.

### 3.3.2 Subjective data

After each story, participants answered a question about the subjective level of effort they had to expend in order to follow and understand the speech. This question was modelled on ACALES (Krueger et al., 2017), an adaptive procedure to determine subjective listening effort. Participants could choose between 14 levels of effort (0 = effortless to 13 = very effortful, 14 = noise only). Results are shown below in Figure 3-2. Overall, lower SNRs were rated as more effortful to listen to than higher SNRs, as for both groups the increase in background noise (from +5dB SNR to 0dB SNR) raised subjective listening effort. In more adverse conditions, subjects also benefited subjectively from the noise reduction algorithm and, to an even greater extent, from the directional microphone mode. Although there were some differences between the groups in specific conditions, no consistent trend was observed. This means that the hearing-impaired group in this study did not report higher levels of effort relative to the normal-hearing group, as often indicated in the literature. This finding could be due to positive effects of the hearing aids regarding intelligibility or listening comfort in the hearing-impaired group, or, conversely, the unusual experience of wearing a hearing aid for the normal-hearing group, which might have raised effort levels for them. Statistical analysis confirmed that SNR level ($\chi^2$ (1) = 331.84, p < 0.0001) as well as type of algorithm ($\chi^2$ (2) = 26.736, p < 0.0001) both had a significant effect on listening effort ratings. There was no significant effect of group ($\chi^2$ (1) =1.3812, p = 0.2399), and no significant interaction effects.

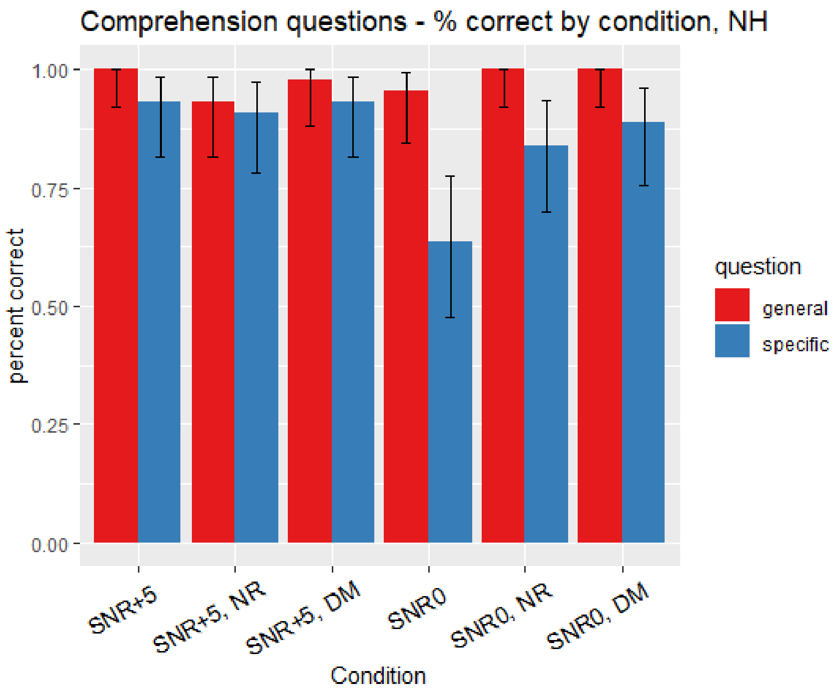
**Figure 3-2: Subjective ratings of listening effort for hearing-impaired and normal-hearing groups, in two SNRs (+5dB, 0dB) and with 3 settings on hearing aids (amplification only, noise reduction, directional microphone)**

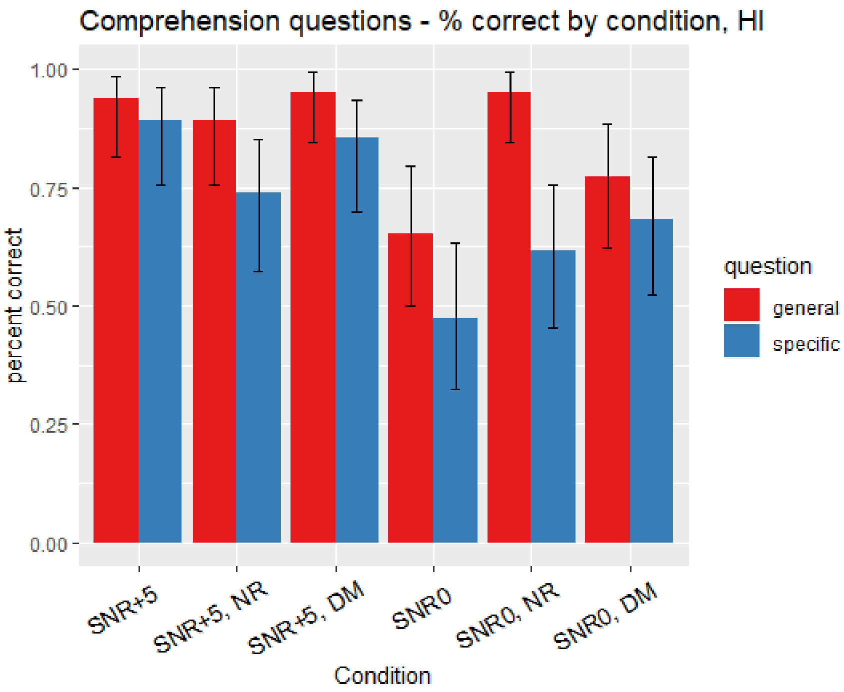


### 3.3.3 Behavioural data

Both groups scored higher on question 1 (general comprehension) than question 2 (detailed comprehension) (see Figure 3-3 below). In the easier conditions (+5dB SNR), both normal-hearing and hearing-impaired subjects performed at ceiling level for the general comprehension question. While performance on the more detailed comprehension question was slightly lower for the HI group, accuracy rates still reached around 75% even for the lowest average performance level. Thus, while the NH group seemed to gain a more detailed understanding of the target speech, intelligibility levels for the +5dB SNR conditions were generally high across both groups. In higher noise levels (0dB SNR), the differences between the groups were more distinct, with the normal-hearing group still performing at ceiling for the general question, whereas the hearing-impaired group reached around 75% correct only. Importantly, in both groups the algorithms (NR and DM) resulted in better speech recognition on detailed questions in high noise levels. In parallel with subjective ratings, directional microphone mode was more effective in raising speech intelligibility than the noise reduction algorithm. For statistical analysis, SNR level, type of algorithm, group as well as question type were considered. All main effects showed statistical significance (SNR level ($\chi^2$ (1) = 26.312, $p < 0.0001$), type of algorithm ($\chi^2$ (2) = 9.8455, $p < 0.01$), group ($\chi^2$ (1) = 43.258, $p < 0.0001$), question ($\chi^2$ (1) = 45.775, $p < 0.0001$)). There was a significant interaction between SNR level and group ($\chi^2$ (1) = 5.1678, $p = 0.023$), SNR level and type of algorithm ($\chi^2$ (2) = 20.611, $p < 0.0001$), and SNR level and question ($\chi^2$ (1) = 6.2606, $p = 0.012$). These results indicate that, while the normal-hearing group outperformed the hearing-impaired group as expected, the hearing-impaired group still had a good understanding of the general topic at least in higher SNR levels. Both groups performed better at answering the general question compared to the more specific one, and both groups benefited from noise reduction and directional microphones for detailed speech understanding. The interactions between SNR levels and all other factors indicates that the differences between the groups, the type of algorithm and the questions is impacted by the amount of background noise. For example, while the algorithms did not aid normal-hearing users in higher SNRs or on general questions overall, they did so in more difficult conditions and with more detailed information on the texts.

**Figure 3-3: Behavioural results of speech comprehension for both groups (top: normal-hearing, bottom: hearing-impaired), in two SNRs (+5dB, 0dB) and with 3 settings on hearing aids (amplification only, noise reduction, directional microphone), with 95%-confidence interval**



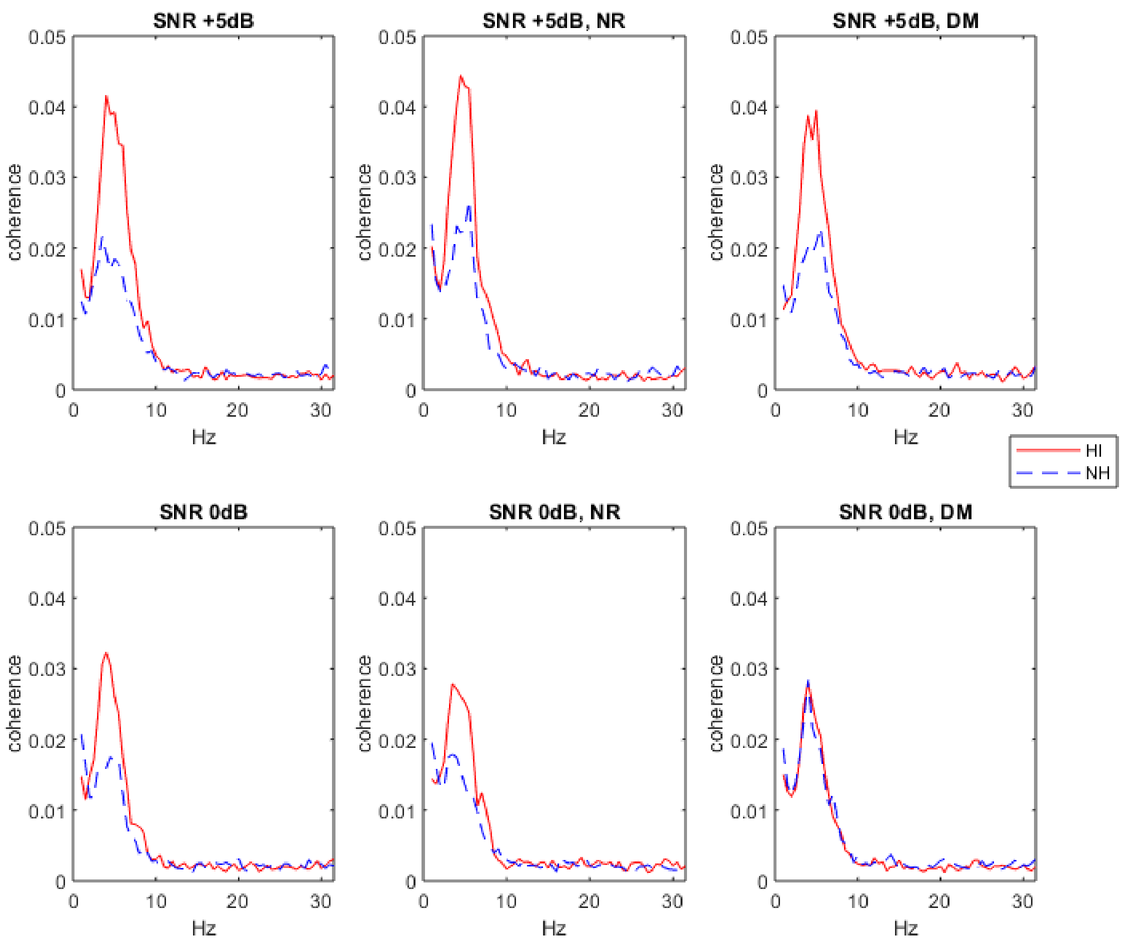



### 3.3.4 Cortical entrainment

Figure 3-4 shows coherence for the normal-hearing (in blue) and the hearing-impaired (in red) groups. Table 3-1 lists mean coherence values for both groups. Most importantly, cortical entrainment for the HI group was vastly increased compared to the NH group, particularly in higher SNRs. For the NH group, entrainment differences between high and lower noise levels were relatively small, and only slightly lower in high noise conditions. These entrainment results from the NH group were highly similar to levels observed in Study 1: while the average coherence in Study 1 for SNRs down to -0.5dB conditions was approximately 0.016, the average coherence in Study 2 across all conditions (i.e. down to 0dB SNR) was 0.017. In the HI group, however, entrainment was markedly decreased in higher noise levels and thus appeared more similar to the NH group: on average, while entrainment for SNRs of +5dB amounted to approximately 0.033, for SNRs of 0dB it was only 0.022. This equals a reduction in entrainment by approximately 35%. In both groups, noise reduction algorithm and directional microphone mode did not impact cortical entrainment in a consistent way. Statistical analysis revealed a significant effect of SNR level ($\chi^2$ (1) = 52.721, $p < 0.0001$), and group ($\chi^2$ (1) = 6.7371, $p < 0.01$), but no significant effect of algorithm ($\chi^2$ (2) = 0.0306, $p = 0.9848$). There was a significant interaction between SNR level and group ($\chi^2$ (1) = 24.645, $p < 0.0001$). The interaction effect between group and SNR levels indicates that the group differences vary depending on the noise level (for the HI group the difference in entrainment between high and low noise levels was much greater than for the NH group).

These findings overall are in agreement with previous research indicating higher entrainment in older, hearing-impaired adults. The consistency of entrainment for the NH group across noise levels suggests that an SNR of 0dB is not high enough to cause entrainment to drop, as was observed in Study 1 for SNRs of -4dB. The failure to record differences in entrainment for noise reduction or directional microphones indicates that cortical entrainment does not capture the same dimensions of effort as subjective and behavioural evidence, which both reported benefits to listeners particularly in more adverse noise conditions. While the algorithms' effects might simply have been too weak to be reflected in cortical entrainment, this measure might also not be the appropriate dimension on which algorithms in hearing aids can be investigated. Since hearing aid algorithms have also been shown to work best in positive SNRs even higher than +5dB (Kortlang et al., 2018), a further increase in SNRs might be necessary to capture the impact on cortical entrainment.

**Figure 3-4: Signal reconstruction performance for cortical entrainment for both groups in two SNRs (+5dB, 0dB) and with 3 settings on hearing aids (amplification only, noise reduction, directional microphone) (top); boxplots and beeswarm plots to show variability of the data (bottom)**



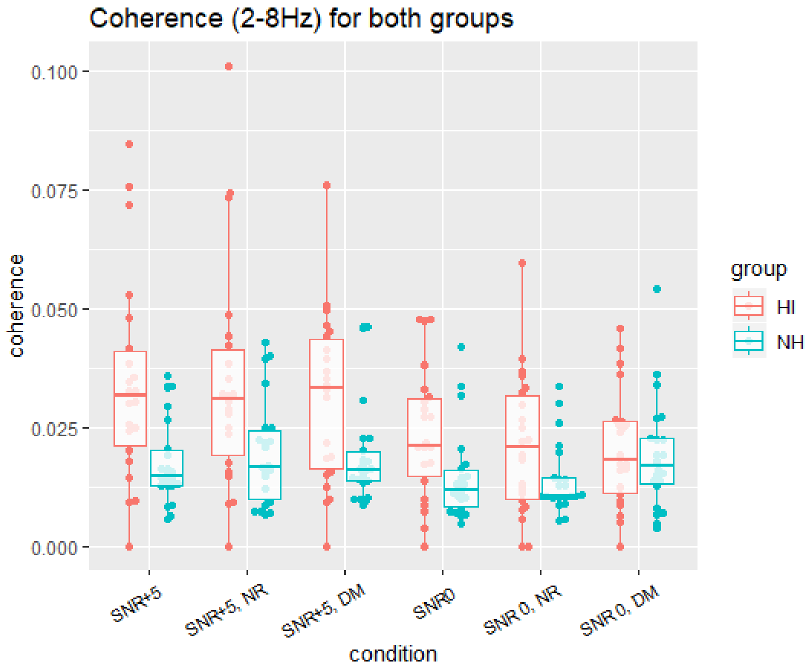


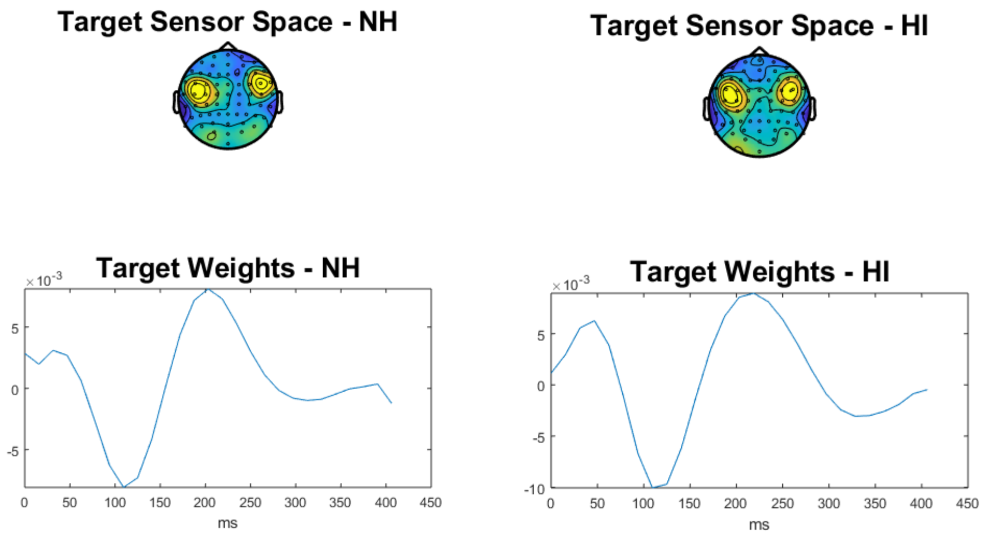

**Table 3-1: Mean coherence values to the target speaker per condition and group**

| Mean coherence value (NH) 2-8Hz | |
|---|---|
| SNR +5dB | 0.0174 |
| SNR +5dB, NR | 0.0195 |
| SNR +5dB, DM | 0.0185 |
| SNR 0dB | 0.0145 |
| SNR 0dB, NR | 0.0141 |
| SNR 0dB, DM | 0.0190 |

| Mean coherence value (HI) 2-8Hz | |
|---|---|
| SNR +5dB | 0.0342 |
| SNR +5dB, NR | 0.0342 |
| SNR +5dB, DM | 0.0310 |
| SNR 0dB | 0.0236 |
| SNR 0dB, NR | 0.0212 |
| SNR 0dB, DM | 0.0206 |

Figure 3-5 shows target sensor space plots and model weights for both groups (best electrode = FC3). Similarly to Study 1, target weights display a shape resembling the P1-N1-P2 component found in ERP studies. These model weights, as well as the sensor space (plotted at the maximum value of around 200ms), seemed to be similar visually for both groups. This might suggest that the normal-hearing and the hearing-impaired group had similar processing of the target speech with regards to latencies, even if entrainment (as shown above) was significantly enhanced for the hearing-impaired group.

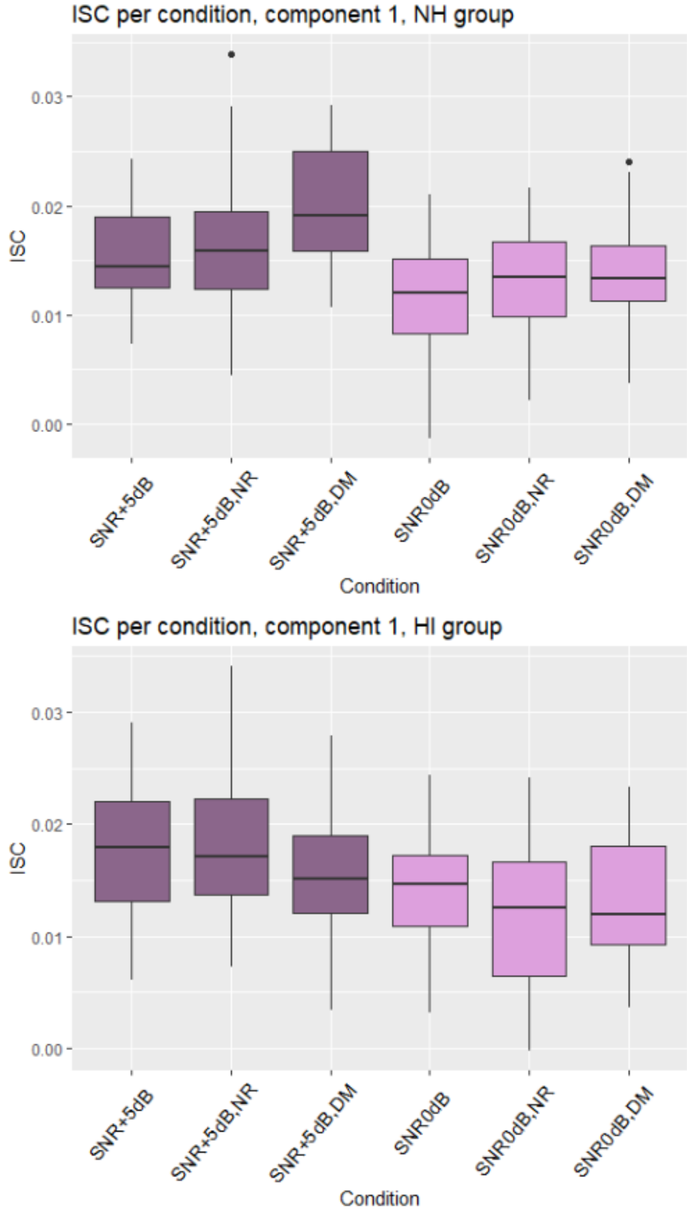
**Figure 3-5: Target sensor space plots (at maximum value) and mTRF model weights for cortical entrainment (left: normal-hearing, right: hearing-impaired)**




### 3.3.5 Intersubject correlation

Figure 3-6 shows intersubject correlation (ISC) per condition for component 1. For both the normal-hearing and the hearing-impaired group, ISC was higher in the easier noise conditions (+5dB SNR) than for the difficult conditions (0dB SNR), which means that the neural response patterns to the stimuli were more consistent across subjects in easier noise conditions. As higher ISC has been associated with higher attention in previous research (Ki et al., 2016), this indicates that both groups were better able to pay attention in higher SNRs. Additionally, the normal-hearing group showed a trend of higher ISC for conditions with enhancement compared to those without. For the hearing-impaired group, this trend was less consistent. Statistical analysis revealed a significant effect of SNR level ($\chi^2$ (1) = 60.566, $p < 0.0001$), but no significant effect of algorithm ($\chi^2$ (2) = 3.5443, $p = 0.17$) or group ($\chi^2$ (1) = 0.0007, $p = 0.9794$). There was, however, a significant interaction between group and algorithm ($\chi^2$ (5) = 17.006, $p < 0.01$). This means that ISC was significantly enhanced in more positive SNRs. However, the two groups were not significantly different from one another in how well they were able to pay attention (as measured through ISC). Furthermore, and similar to results from cortical entrainment, the noise reduction algorithm or directional microphone mode did not result in higher levels of attention. Reasons for the failure to record such effects might be similar to those for cortical entrainment; ISC might either not be an appropriate dimension along which to evaluate the impact of hearing aid algorithms, or the SNRs tested in this study might not be at the right levels to show a significant difference. The interaction between group and algorithm, however, suggests that the HI group was impacted by the algorithms in different ways from the NH group and confirms the visual interpretation that ISC in conditions with algorithms was enhanced in the NH group but not in the HI group.

**Figure 3-6: ISC per condition, component 1 for both groups (top: normal-hearing, bottom: hearing-impaired) and split by condition**




## 3.4 Discussion

This study focused on differences in measures of auditory attention and processing between a normal-hearing and a hearing-impaired group at different levels of SNR, as well as the effects of different algorithms in hearing aids on those measures. Differences between the normal-hearing and the hearing-impaired groups were shown behaviourally (in higher noise levels and with more detailed information, the HI group showed lower levels of speech intelligibility) and on cortical entrainment, where the HI group had substantially increased entrainment compared to the NH group, particularly in lower noise levels. While all measures (subjective, behavioural and neurophysiological) were impacted significantly by SNR level, the addition of noise reduction or directional microphone mode was only beneficial in significant ways for subjective ratings and speech comprehension scores, particularly in higher noise levels.

### 3.4.1 Auditory attention in older, hearing-impaired people

Based on findings from previous studies on hearing impairment (Dubno et al., 1984; Helfer & Freyman, 2005; Hogan & Turner, 1998), it was assumed that hearing-impaired people would have more difficulties understanding speech in multi-talker babble noise. Behavioural results, functioning as a proxy for intelligibility, indicated that the NH group did indeed outperform the HI group, particularly on the more specific comprehension questions and in higher noise levels. Sensorineural hearing loss, resulting from damage to hair cells in the cochlea, degrades the acoustic representation of incoming sound (Bernstein & Oxenham, 2006; Lorenzi et al., 2006; Moore et al., 1999), which is at least partly responsible for worse speech perception performance in adverse conditions in hearing impaired people. The present study sought to ameliorate these differences in bottom-up processing by providing the HI group with individually fitted levels of sound amplification through their hearing aid. Moreover, NH subjects were fitted with a hearing aid with minimal amplification, in order to replicate the listening experience HI subjects had. While in lower noise levels and for more general speech comprehension, HI listeners seemed to indeed profit from their hearing aids, in higher noise levels they were less able to do so.

It has been shown that older adults, perhaps in order to compensate for reduced perceptual abilities in noise, rely more heavily on context to understand speech than young adults do (Divenyi, 2005). A study with hearing aid users has also found that semantic context improves speech intelligibility and reduces listening effort in hearing-impaired subjects (Holmes, Folkeard, Johnsrude, & Scollie, 2018). These effects of age and hearing impairment might explain why HI subjects in the present study still performed relatively well on general speech comprehension. As this study used stories rather than just sentences in isolation as stimuli, HI subjects might have been able to rely on contextual information in order to answer comprehension questions about the texts heard.

While it might be expected that cortical entrainment would correlate with intelligibility levels (Vanthornhout, Decruy, Wouters, Simon, & Francart, 2018), and thus would tend to be at least slightly lower in the HI group, the opposite effect was observed in this study. Entrainment in the HI group (for SNR levels of +5dB) were approximately twice as high as for the NH group. In the present study, there are two potentially influential, confounding factors on envelope tracking for the NH and the HI group: age and hearing loss. Since the NH group had a lower average age (37.9) than the HI group (67.5), and since ageing as well as hearing impairment have been shown to impact entrainment, both factors needed to be considered. Although not many studies so far have been conducted on age, hearing impairment, and their respective influence on neural entrainment, existing research shows similar results as those found in the present study, indicating that age and hearing impairment both seem to increase rather than decrease neural entrainment (Goossens et al., 2018; Millman et al., 2017; Presacco et al., 2016b, 2016a).

Factors that might contribute to the greatly enhanced cortical entrainment for the older, HI group are potential peripheral as well as central changes in language processing. Magnified envelope coding might be indicative of a processing imbalance of temporal cues, through which envelope cues are processed to an enhanced degree while temporal fine structure cues are processed less than in young, normal-hearing people due to impaired frequency selectivity and the loss of the active mechanism in the cochlea (Moore et al., 1996; Sek et al., 2015). Since it has been shown that the enhanced availability of envelope cues impairs speech perception in normal-hearing people and in particular in noise (Moore & Glasberg, 1993), a similar mechanism might be active in older, hearing-impaired people with greatly increased envelope tracking despite lower levels of speech perception. In addition, previous research has indicated that both age and hearing impairment also have adverse effects on central processing mechanisms involved in selective auditory attention (Dai et al., 2018; Helfer & Freyman, 2008; Holmes et al., 2017; Tun et al., 2002). fMRI studies have further shown that ageing reduces the activation of, and connectivity between, certain language areas in the brain (Peelle et al., 2010). Thus, enhanced envelope tracking and worse speech perception performance might result from both peripheral and central changes in neural speech processing that are associated with age and hearing impairment.

Apart from these considerations concerning processing of the target, the suppression of distractor sounds (Fiedler, Woestmann, Herbst, & Obleser, 2019) might be less effective in adverse conditions in HI listeners compared to NH listeners. In sum, these findings can be interpreted as an indicator that older and hearing-impaired adults may use different strategies to process speech in adverse conditions, due to less effective inter-regional connectivity between brain areas, imbalances related to the processing of temporal characteristics of speech, and potentially weakened mechanisms to suppress distractor sounds.

Aggregating previous findings, a complicated picture emerges of the relationship between intelligibility, ageing, hearing impairment and neural envelope tracking. On the one hand, intelligibility and performance on speech tasks have generally been associated with enhanced neural tracking (Ahissar et al., 2001; Peelle & Davis, 2012), thus linking higher entrainment with positive effects: between-subjects analyses of studies with the attentional selection entrainment effect have shown that subjects with higher entrainment generally performed better on speech recognition tasks (O'Sullivan et al., 2015). On the other hand, ageing and hearing loss have equally been associated with higher neural entrainment (Goossens et al., 2018; Millman et al., 2017). In brief, while cortical entrainment is associated with positive characteristics (more attention, better intelligibility, higher performance) in young, normal-hearing adults, this relationship appears to be change with age and hearing impairment. Compared to a healthy control group, older and hearing-impaired people do not generally benefit from enhanced envelope tracking. Rather, this phenomenon might be generated by the reduced ability to process temporal fine structure and a resulting greater focus on envelope cues.

For normal-hearing subjects, neural entrainment remained stable across SNR levels (+5dB and 0 dB). This finding concurs with those from Study 1 and other previous studies (Ding & Simon, 2013), which suggest that neural entrainment to an attended speaker in noise remains relatively stable for SNRs as low as 0dB. While the present study differs from the experimental design in Study 1 with respect to the stimuli used (sentences vs. stories) and, more importantly, the additional intelligible distractor speaker present in Study 1, the results from cortical entrainment in easy to moderate noise conditions are strikingly similar between these two studies. One inference that can be drawn from these similar results is that one of the main factors impacting the extent of target talker entrainment is noise, as this was present (and at similar levels) across the compared studies and conditions.

In contrast to the normal-hearing subjects, for which entrainment across different SNR levels was relatively stable, higher noise levels (SNR 0dB) led to a drop in entrainment for the hearing-impaired group. It could be speculated that this fall in entrainment might be an effect resembling that observed in NH subjects at very high levels of noise (such as in Study 1 at an SNR of -4dB). While the SNR of 0dB is considerably higher than -4dB in Study 1, it is possible that hearing-impaired subjects simply exhibit a lower level of noise tolerance before entrainment drops off, compared to normal-hearing subjects. However, behavioural performance for the -4dB SNR condition in Study 1 was relatively low with an average of well below 25%, whereas performance for the hearing-impaired group in Study 2 for SNR levels of 0dB reached almost 70%. Consequently, while in normal-hearing adults, the decrease in entrainment at high noise levels might be associated with a concurrent drop in intelligibility, it is difficult to infer the same for hearing-impaired adults, based on the vastly different levels of speech perception.

Analyses of ISC revealed that on the whole, ISC was higher for the easier noise conditions (+5dB SNR) in both groups, but there were no significant differences between the groups or between conditions with added signal processing (NR or DM). While the literature on ageing, hearing impairment and neural entrainment is scarce, there are even fewer studies investigating the relationship between age, hearing impairment and ISC. Mixed findings have been presented on the effect of age on ISC (Campbell et al., 2015; Cantlon & Li, 2013), and hearing impairment has never before been studied in conjunction with ISC. In this study, importantly, levels of ISC were not significantly different between the two groups, which stands in contrast to findings from cortical entrainment. While both cortical entrainment and ISC have been used as measures to study auditory attention in adverse listening situations, they might refer to different aspects of attention. Furthermore, the lack of ISC differences between the groups relates to the subjective measure observed in this study, which also did not find any differences between the groups. One conclusion that can be drawn from the combination of these results is that the HI group might not consciously have paid more attention, and hence also did not subjectively perceive to have put in more effort. The group differences observed for cortical entrainment might rather be attributable to unconscious processes in the HI group such as the less efficient suppression of the distractor or reduced activation and interconnectivity between certain language areas in the brain (Peelle et al., 2010).

### 3.4.2 Impact of HA algorithms on neurophysiological processing

Findings about the noise reduction (NR) and directional microphone (DM) mode indicate that, while they significantly reduced subjective ratings of listening effort and increased speech perception in both groups, no effects were observed on cortical entrainment or ISC. One explanation for the failure to find effects of processing algorithms on neurophysiological measures might be that the effect size was too small, compared to the vast effects of hearing impairment and age on neural entrainment. Another possibility is that these measures do not capture the dimension along which hearing aid users experience the benefits; while they were clearly observed subjectively and behaviourally, they might not specifically impact neural processing in ways measurable by cortical entrainment or ISC.

In the present as well as in previous studies (Alcántara et al., 2003; Boymans & Dreschler, 2000; Picou, Moore, & Ricketts, 2017), listeners have been shown to subjectively prefer modes such as noise reduction or beam forming in hearing aids. At the same time, it has been notoriously difficult to connect these consistent subjective results with objective measures. In this regard, findings from this study are well in line with previous research. The questions that thus remain are whether and how listening effort can be captured with an objective measure, and if so, whether the neurophysiological measures employed in this study do not assess the appropriate dimensions, or whether they are simply not sensitive enough to capture differences in listening effort between conditions with and without signal processing algorithms in hearing aids.

To sum up, this study reported differences between normal-hearing and hearing-impaired people on measures of neural speech processing. Much higher cortical entrainment for the hearing-impaired group was observed, which might be explained by different neural processing strategies in older, hearing-impaired people compared to younger, normal-hearing groups. The observation that general measures of intelligibility remained relatively high for the hearing-impaired group could be due to their reliance on context to aid speech perception in noise. Intersubject correlation as a global measure of attention did not differ between the groups but across SNR levels, indicating that both groups could more easily pay attention to the target stream in lower noise levels. While noise reduction and directional microphone modes lowered subjective listening effort and facilitated more detailed speech perception, this impact was not reflected in neurophysiological measures such as cortical entrainment or ISC. Potentially, the effects were too small to be significant, or the measures did not capture the right dimensions of processing.

# Chapter 4: The relationship between motivation, task demand, effort, and neural speech processing

## 4.1 Introduction

The first two studies of this thesis have investigated the effect of different types and levels of degradation on different populations. The assumption underlying both these studies was that subjective, behavioural and neurophysiological measures of speech processing vary as a function of task difficulty and task type, and that this variation allows us to gain insights into the working mechanisms of auditory attention and effort. However, there are certain factors that make the interpretation of results not as straightforward as it may seem at first sight. Concerning at least subjective and behavioural measures, there is evidence to suggest that the relationship between task demand and performance is not direct and linear, but mediated by other factors such as arousal; it has been shown that increased arousal can help improve performance, whereas excessive arousal hampers performance.

While in Studies 1 and 2, there is some evidence to suggest that even neurophysiological measures of speech processing might reflect some form of a non-linear relationship between task demand and performance (e.g. cortical entrainment to a target speaker in noise remains relatively stable in low to moderate noise conditions, but drops away when noise levels become too high), it was not the main question of interest. The present study, in contrast, specifically focused on this observed relationship between task demand and performance. This study aimed to further explore the link between top-down, central cognitive mechanisms and neurophysiological measures of language processing. Specifically, an incentive for subjects to perform better was manipulated with the intention to alter their motivational state. By doing so, it was expected that different levels of motivation would influence effort levels, behavioural performance and potentially even neurophysiological measures from our toolbox. By explicitly manipulating motivation for the task, this study investigated the influence of top-down factors on entrainment, which is currently still a topic of considerable debate (Ding & Simon, 2014).

On top of the impact of motivation, this study investigated potential factors underlying increased cortical entrainment observed in Study 2 for older, hearing-impaired people through a specific type of signal degradation. This type of degradation decreased intelligibility and spectral detail, but kept periodicity and the broadband amplitude envelope intact. It was thus explored whether the reduced spectral information available would result in higher entrainment previously only recoded in older, hearing-impaired or L2 speakers. In summary, this study examined both the specific effects of changing effort (through the manipulation of motivation) and the broader implications that can be drawn from this experimental manipulation about top-down features and their impact on neurophysiological measures, and potential reasons for increased cortical entrainment despite lower speech intelligibility levels.

### 4.1.1 Motivation and task performance

One of the earliest descriptions of the relationship between effort and task difficulty dates back to Kukla (1972), who posited that effort (described there as the 'vigor with which a task is undertaken') steadily increases with task difficulty, but drops again when the task becomes too difficult. Among the factors that influence effort, Kukla listed the following: the task's perceived difficulty, the subject's perception of their own abilities, and the experience of success and failure. Notably, all these factors highlight the subject's perception and experience, rather than objective measurement, of their own capabilities and task difficulty. This underscores the importance of prediction and anticipation with regards to the mental capabilities that are going to be deployed for any specific task at hand.

The Motivational Intensity Theory, introduced by Brehm & Self (1989), further refined the relationship between task demand and effort. This theory posited that the allocation of effort is centrally dependent on the resource/energy conservation principle: in order to minimize the waste of resources, the energy invested in a task should equate, but not exceed, what is needed to complete it successfully. In other words, optimal effort investment is matched to the effort required for task completion. However, Motivational Intensity Theory, similarly to Kukla, also proposes that an important requirement for any kind of effort investment is the expectation that task success is (subjectively) possible and the required effort justified. Motivational arousal increases with task difficulty, but only up to the point where either effort exceeds justification, or the task exceeds the person's capabilities (i.e. success is not possible). Evidence for the Motivational Intensity Theory has been introduced in the form of physiological measures, specifically cardiovascular reactivity (Gendolla & Wright, 2005; Gendolla, Wright, & Richter, 2012; Wright, 1996). These studies have shown that subjective task difficulty (perceived chances of success) as well as motivation (e.g.

material incentives, prospect of being evaluated on the test material, etc.) do indeed influence effort (at least when measured through cardiovascular reactivity). Furthermore, subjectively reported reduced motivation has also been associated with a decrease in mean amplitude of the ERP measure N1, an index of general arousal (Moore, Key, Thelen, & Hornsby, 2017). All these studies substantiate the connections between motivation, arousal and effort, and provide evidence that these connections can indeed be captured with objective measures.

Particular groups of people that are more vulnerable to adverse listening conditions might require more incentives (i.e. greater motivation) to sustain the same levels of performance. For example, one study with children with and without ADHD (Attention Deficit Hyperactivity Disorder) has shown that those with ADHD experienced higher cognitive load in difficult listening conditions, and that they also required more motivation (in the form of feedback) than their peers without ADHD (Russell, 2015). More generally, it has been found that motivation plays a significant role in treatment adherence for chronic health conditions (Vermeire, Hearnshaw, Van Royen, & Denekens, 2001). A study on computer-based auditory training to improve speech-in-noise perception for hearing-impaired people has reported that motivation is a decisive factor for keeping people engaged in the program (Henshaw, Mccormack, & Ferguson, 2015). These studies stress the importance of motivation as an important predictor of treatment outcomes for vulnerable or impaired populations. Therefore, the measurement and manipulation of motivation is a central area of research when trying to decode, understand and optimize human auditory attention in challenging situations.

### 4.1.2 Listening effort and motivation

Listening effort has been defined in various ways, with different research purposes in mind. One useful definition used throughout this thesis is "the mental exertion required to attend to, and understand, an auditory message" (McGarrigle et al., 2014). While this definition stresses the central link between listening effort and attention, it misses the important element of motivation that this study focuses on. Another definition, taking into account the importance of the listener's aims and motivation, describes listening effort as the "deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task that involves listening" (Pichora-Fuller et al., 2016). It emphasises the fact that effort allocation is a deliberate, rather than automatic process, which can be flexibly adapted depending on the task at hand and its importance to the listener. While an adverse listening situation in a noisy environment does not automatically imply high listening effort in and by itself, it does so if it is of importance to the listener to follow and understand a specific conversation.

While some studies and theories have loosely made links between motivation, effort, and auditory attention, not many have investigated these links explicitly. However, this connection is particularly important with a view to the importance of intent in listening and understanding an auditory message as defined in the Framework of Effortful Listening (FUEL, Pichora-Fuller et al., 2016). While most studies on listening effort have manipulated task difficulty (e.g. level or type of noise), and then attributed the change in a neurophysiological measure (e.g. increase in pupil size) to listening effort, it has been argued that a good measure of listening effort should also be validated by showing that it varies depending on other factors, such as motivation (Richter, 2016).

In an experiment with subjective ratings of listening effort, it has been found that motivation increases measures of both listening effort and tiredness, likely because higher motivation increases efforts to complete the task (Picou & Ricketts, 2014). In this study, motivation manipulation was operationalized by asking subjects to either listen to the speech only (low motivation) or listen to the speech and then answer quiz questions to test their understanding (high motivation).

The only study so far examining motivation in conjunction with an auditory task and physiological measures (cardiovascular activity) corroborates many of the predictions generated from the motivational intensity theory. In this study (Richter, 2016), success importance (motivation) was operationalized through a performance-dependent monetary reward system. In addition, task difficulty was varied by presenting subjects with sine tones that were highly similar (difficult discrimination condition) and highly distinct (easy discrimination condition). It was found that, in line with predictions generated from the Motivational Intensity Theory, cardiovascular reactivity was low in easy conditions. In difficult discrimination conditions, cardiovascular reactivity depended on reward (high reactivity if success was important, low reactivity when success was not important). This experiment, although not conducted with continuous speech material, demonstrates the link between motivation, effort, and attention, and how these are reflected differently in a neurophysiological measure. Based on the acknowledgment that motivation is an important factor determining successful speech recognition in adverse conditions, it has also been argued in theoretical papers on listening effort that one possible intervention to reduce listening effort is to change the subject's perception of the task demand by changing motivation and the perception of self-efficacy (the perception that one has the abilities and capabilities to achieve their aims) (Pichora-Fuller, 2016).

An important aspect to consider when manipulating motivation is the question of how motivation can be best operationalized. In the two studies presented above, motivation was manipulated by either the threat of testing subjects on their knowledge after listening to stimuli, or by varying the monetary reward depending on performance. However, there are downsides to both of these manipulations: the 'threat' of a test on the subject's knowledge fails to take into consideration that many participants in scientific studies take part primarily for monetary reasons, and might not be significantly more motivated by knowing they might be tested on what they had heard, and that they might perform poorly when not paying attention as best as possible. Varying the monetary reward, at the same time, is arguably a highly subjective motivational factor, as one subject might be very keen to increase their earnings a bit, while another subject that is not short on financial resources might not be that motivated by the prospect of higher monetary reward. For these reasons, another strategy was used to manipulate motivation in the present study. The incentive for subjects to perform better came in the form of saving time: in incentive conditions, subjects could shorten the time they had to spend doing the auditory task through better performance. For non-incentive conditions, in contrast, subjects had no such performance-dependent motivator to do better. Participants were offered a fixed fee regardless of how long they spent doing the task.

### 4.1.3 Signal degradation and increased cortical entrainment

Results presented in Study 2 as well as previous research have demonstrated that cortical entrainment in certain populations (older, hearing-impaired or L2 listeners) can be greatly increased compared to their normal-hearing or native listener peers (e.g. Presacco et al., 2016b; Song & Iverson, 2018). While there is some indication that neural processing might change with age and/or hearing impairment (Peelle et al., 2010), it is unclear why these populations show much higher cortical entrainment to speech despite lower levels of speech intelligibility.

One reason for this increased entrainment might be found in the ability to process spectral information and modulations contained in speech. There is behavioural evidence that hearing-impaired people benefit less than normal-hearing people from additional temporal fine structure information in adverse conditions, while at the same time relying more on envelope cues (Hopkins et al., 2008; Sek et al., 2015; Wallaert et al., 2017). Amplitude and frequency modulations contained in speech have both been demonstrated to be important to speech intelligibility in silence and with background talkers (Varnet, Ortiz-Barajas, Guevara Erra, Gervain, & Lorenzi, 2017), and cortical entrainment might reflect the extent of their availability. The seemingly contradictory results between hearing-impaired, older people and normal-hearing, younger people could be reconciled by demonstrating experimentally that increased entrainment can also be triggered in normal-hearing people with a degradation type that both preserves the temporal broadband envelope and periodicity information, while at the same time reducing the spectral detail. This hypothesis was tested in the present study by integrating a type of degradation that fulfils these criteria into the experiment alongside the factor of incentive as outlined above. Speech intelligibility levels were adapted individually to create two levels of degradation to be compared with a non-degraded condition.

The present study is the first to investigate the effect of motivation on a variety of neurophysiological measures, as well as the impact of a specific signal degradation with reduced spectral detail but retained periodicity. Subjects were presented with a listening task (1 target speaker and 2 distractors), and degradation was added to the target to investigate the form and intelligibility of the speech as well as its representation at different neural processing stages. This study investigated the following questions: What effect does the manipulation of an incentive to perform better have on neurophysiological measures? What can these results reveal more broadly about the influence of top-down determinants of neurophysiological measures, and in particular cortical entrainment? Can a specific type of degradation activate a mechanism of increased entrainment in normal-hearing people previously only recoded in hearing-impaired, older and L2 listeners?

## 4.2 Methods

### 4.2.1 Subjects

Twenty-one adults (m = 9, f = 12; mean age = 23.14) participated in the study. All subjects were native British English speakers. All subjects were healthy and had self-reported normal hearing. Subjects were paid a fixed fee, and the experimental procedure was approved by the UCL institutional ethics committee. Written informed consent was obtained before the experiment.

### 4.2.2 Stimuli

The same sentence material as in Study 1 was used, with varying levels of predictability: for this study a total number of 987 sentences were used, spoken by a female Southern British English native speaker. Of these, 439 sentences each (44.5%) were of high and low cloze predictability, respectively. The remaining 109 sentences (11% in total) were anomalous ones. As distractors, stories recorded with the same female British English native speaker were used.

To distort the target speaker, a particular type of harmonic distortion was added that flattened out details of the spectrum but preserved f0 and the broad-band amplitude envelope. The distortion was computed by first extracting the analytical waveform through the following method: An analytical fundamental waveform was computed from the speech materials through the following method: first, the signal was filtered to extract the first harmonic. This was done by measuring the fundamental frequency (f0) contour using the program 'fxrapt' as implemented in Matlab (Talkin, 1995), dividing the signal into 250ms windows, and using a zero-phase low-pass Butterworth filter of the $5^{th}$ order in each segment set to 50% greater than the maximum f0 in the window. The envelope of this first harmonic was remodulated so that it matched the low-pass broadband envelope of the signal. This was done by calculating the Hilbert envelopes of the low-pass filtered (50 Hz) original signal, and adjusting the envelope of the first harmonic to match these. To create the distortion, the analytical waveform was converted to a buzz that matched the long-term spectrum of the original speech signal. This was done by finding the zero crossings from low to high in the f0 signal, putting a biphasic pulse at each of those points, then filtering it. This filter was constructed by calculating the average spectrum of the original talker, and matching the pulse train to this spectrum. Then, the envelope was remodulated so that it matched the broadband amplitude envelope of the original signal.

This distortion type was different from those used in previous experiments in various ways: compared to speech-shaped noise, it retained the information contained in the amplitude envelope (rather than just matching the long-term average spectrum). Compared to a vocoder, f0 was preserved rather than eliminated.

At the beginning of the experiment, participants' individual speech reception threshold (SRT) levels were determined through an adaptive procedure. Based on the results from this procedure, two levels of the distortion as described above were added to the target speech: one at the SNR of their individual SRT, and one where the SNR was raised by 5dB compared to the threshold level. Together with the presence or absence of an incentive, the following six conditions were used:

1) No incentive, no distortion

2) No incentive, above threshold (SRT + 5dB)

3) No incentive, threshold (SRT)

4) Incentive, no distortion

5) Incentive, above threshold (SRT + 5dB)

6) Incentive, threshold (SRT)

### 4.2.3 Apparatus

An EEG experiment was conducted (Biosemi Active Two system with 64 (Ag/AgCl) electrodes) with 7 reference electrodes placed on the face to control for eye blinks and movement as well as muscle activation-induced artefacts. Participants were seated in a sound-attenuated booth. The experimental task took approximately one hour to complete.

### 4.2.4 Procedure

Subjects were seated in a sound-attenuated booth. Before starting the main experiment, they completed a short adaptive procedure to determine their individual speech reception thresholds (SRT). During the adaptive procedure, subjects listened to sentences by the female target speaker with varying levels of degradation, with the additional presence of two distractor voices located at 45 degrees to the right and 45 degrees to the left. These distractor recordings were also spoken by the target speaker, as in the main experiment. After the presentation of each sentence, subjects were asked to repeat what they had heard. These responses were scored in a binary fashion (either all content words correct or not) for 20 sentences in total. Based on these results, individual SRTs were determined for subsequent use in the main experiment. Across all subjects, the average SRT was at a signal-to-noise ratio (SNR) of -1.43dB. Apart from the determination of SRTs, the adaptive procedure also familiarized participants with the presentation mode of the main EEG experiment, as target and distractor voices as well as locations were the same.

Following the adaptive procedure, participants listened to 12 stimuli blocks, again with sentence spoken by the target from the front and stories by two distractor voices (using the same talker as the target) located at 45 degrees to the right and 45 degrees to the left, respectively. For 6 blocks ('non-incentive blocks'), the duration was fixed to approximately 3.5 minutes, whereas the remaining 6 blocks ('incentive blocks') were considerably longer (approximately 5.5 minutes) but could be completed faster with good performance. The task for participants was to press a button whenever they heard an anomalous sentence.

The button presses triggered two different feedback loops: for non-incentive blocks, subjects saw either a '+' or a '-' sign on a small screen in front of them, depending on whether they pressed the button correctly or not. However, their performance did not impact the length of the block. For incentive blocks, the screen displayed a counter (starting with 0 at the beginning of the block). For every correct answer, subjects received 3 points, and for every incorrect answer, one point was deducted from their total score. As soon as subjects reached a certain number of points, the block stopped and they could move on to the next one. Thus, the better they performed on incentive blocks, the faster they could be done with the experiment.

Incentive and non-incentive blocks were presented in alternating fashion. Before each block, subjects were told whether it would be an incentive block or not, and if so, how many points they needed to gain in order to finish this block. For blocks where degradation was at threshold level (the most difficult condition), subjects needed 15 points (i.e. 5 correct responses), and for above-threshold as well as non-degraded conditions the number of points was 27 (i.e. 9 correct responses). The experiment was designed so that with perfect performance, subjects could finish an incentive block in the same amount of time as the non-incentive blocks (3.5 min). The maximum time of incentive blocks was approximately 5.5 min. Beyond the 3.5 min threshold and towards the end of incentive blocks, anomalous sentences would occur more frequently so that subjects could reach the required number of points more easily.

### 4.2.5 EEG analyses

An EEG experiment was conducted (Biosemi Active Two system with 64 (Ag/AgCl) electrodes) with 7 reference electrodes placed on the face to control for eye blinks and movement as well as muscle activation-induced artefacts. Participants were seated in a sound-attenuated booth. The experimental task took approximately one hour to complete. Data were recorded at 2,048 Hz. Two electrodes placed on the mastoids were used as reference. Noisy channels were interpolated. For brainstem analyses, a high-pass filter of 0.1 Hz was applied. For cortical analyses, data were band-pass filtered between 0.1 and 30 Hz and downsampled to 256 Hz. All filters used were zero-phase causal Butterworth filters of the 3$^{rd}$ order (response slope of -18dB per octave) as implemented in the ERPlab toolbox (Lopez-Calderon & Luck, 2014) of EEGlab (Delorme & Makeig, 2004). Artefact rejection was performed automatically on all trials, with a rejection threshold of ±150 μV. Independent Component Analysis was applied manually and components associated with eye movements as well as eye blinks with highly spatially stereotyped scalp projections (Onton & Makeig, 2006) were removed from the data. All pre-processing procedures were performed in Matlab using the Fieldtrip toolbox (Oostenveld et al., 2011).

**Frequency following response.** EEG data were processed at the original sampling rate of 2,048 Hz and passed through a Butterworth filter between 100 and 300 Hz. The EEG data were matched with the analytical waveform of each block. The analytical waveform was computed in the same way as in Study 1: first, the fundamental frequency (f0) of the stimulus was calculated and mapped onto a contour in synchrony with the original stimulus. Then, the stimulus was divided into 250ms-long windows. Each window was multiplied by a Hanning-window, and adaptively low-pass filtered at the f0. The waveform was then recomputed to produce a signal matching the original signal with regards to f0 and phase. To concentrate on the amount of periodicity in the signal, it was modulated by the amplitude envelope of the low-pass (50 Hz) original signal.

To the EEG data, a Butterworth filter between 100 and 300 Hz was applied to concentrate on frequency regions around the FFR. Model weights and constants for time lags from -15ms to 45ms were computed, and the backward model from the mTRF Toolbox (Crosse et al., 2016) was trained for the target, reconstructing (predicting) the analytical waveform for each block based on the 11 other blocks. Correlation values between the reconstructed and the actual stimulus were obtained to calculate FFR tracking by condition.

**Cortical entrainment.** EEG data were downsampled to 256 Hz and passed through a Butterworth filter between 0.1 and 30 Hz. The EEG data were matched with each respective amplitude envelope of the stimulus block. As in Studies 1 and 2, the backward model from the mTRF Toolbox (Crosse et al., 2016) was trained for both target and distractors (using time lags of 0ms to 400ms), predicting the amplitude envelope of each stimulus block (1 out of 12) based on the model generated from the other 11. Coherence values were computed to calculate cortical entrainment for each condition separately. This was done by segmenting the data in to 1-sec Hann windows with 50% overlap, and calculating coherence from the cross-spectral density of the FFT of the two signals, divided by the power spectrum of each signal.

**Intersubject correlation.** For intersubject correlation (ISC), data were downsampled to 64 Hz. Intersubject correlation was calculated in the same way as for Study 2, using the toolbox provided by (Cohen & Parra, 2016): the data were concatenated for each subject and compiled across subjects. The cross-covariance between subjects as well as within- and between-subject covariances were computed. Correlated components were calculated and sorted with regards to intersubject correlation strength. ISC for the first component was computed for each condition separately.

**N400.** Data were downsampled to 256 Hz. The data was segmented into epochs time-locked to the final-word onsets (200ms pre-stimulus and 800ms post-stimulus intervals). Trials with amplitude exceeding ±150 μV were automatically rejected. A non-parametric, cluster-based permutation analysis was performed on the EEG data over a time window of 200ms to 600ms after the onset of the last word in order to test for significant clusters of both time and sensor space (Maris & Oostenveld, 2007). This analysis avoids the problem of multiple comparisons (i.e. the problem that there is an extremely large number of time and sensor pairs) by creating clusters in neighbouring time points and electrodes. Significant differences between two conditions (e.g. high cloze vs. low cloze probability sentences) were calculated with Monte Carlo simulations. Specifically, trials were randomly assigned to each condition, and 1,000 of these random partitions were generated. Statistical significance was determined by calculating the proportion of partitions with a higher test statistic than the observed data (summed t-values in a cluster), called the p-value. If this p-value was smaller than the critical alpha-level (0.05 in this case), it was concluded that the data in the two conditions (e.g. high vs. low cloze probability sentences) were significantly different. Despite its advantage of not requiring any a priori selection of electrodes, this cluster-based permutation analysis is difficult to use when investigating interaction effects involving more than two variables. For this reason, for main statistical analysis the N400 amplitude was quantified by averaging the response to high and vs. low predictability sentences across a 300ms to 500ms window across five midline electrodes (Fz, FCz, Cz, CPz, and Pz), following previous N400 studies (Song & Iverson, 2018; Strauß, Kotz, & Obleser, 2013) and Study 1 of this thesis. Prior to statistical analysis, the average responses for each participant were visually inspected, and 2 participants were excluded from further analysis due to the amount of noise in the data. Specifically, the average responses for these participants had amplitude ranges beyond the chosen cut-off amplitude of ±10 μV.

## 4.3 Results

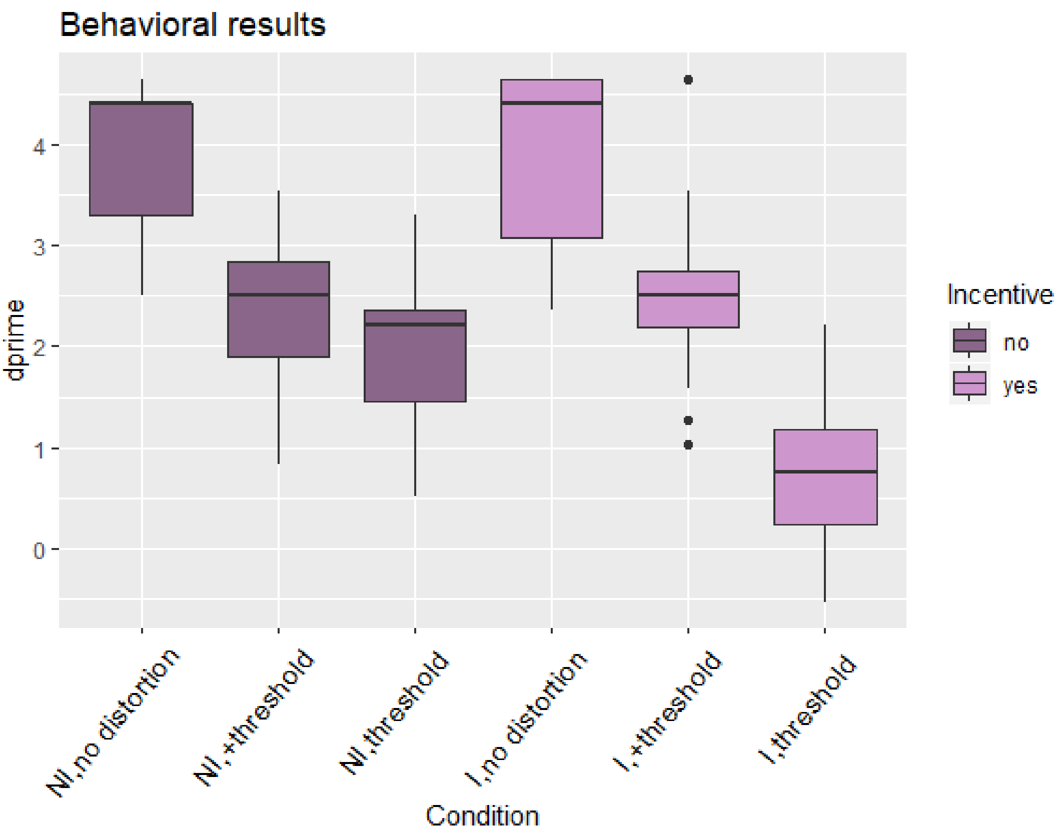
### 4.3.1 Statistical analysis

All statistical analyses were conducted in RStudio version 1.1 (R Development Core Team, 2013). Linear mixed-effects models were applied for statistical analysis, using the lme4 package (Bates et al., 2014). Signal degradation and incentive were considered as main factors and subject was treated as a random factor.

### 4.3.2 Behavioural data

Subjects' button presses were recorded as a behavioural measure of performance. Figure 4-1 below shows speech perception levels per condition across all subjects. With increasing degradation, speech perception performance dropped in both incentive and no-incentive conditions. There was no difference in performance between incentive and non-incentive conditions, except for the hardest (threshold) conditions in which participants performed worse with incentive. This indicates that, overall, the incentive did not lead to an increase in performance. However, in the most difficult condition where subjects presumably perceived their chances of finishing early as very low (incentive condition), they gave up. In contrast, when their performance did not affect the length of the block (non-incentive condition), subjects kept trying and thus received a higher score than in the incentive condition. Statistical analyses with model comparison revealed that signal degradation had a significant effect on d' ($\chi^2$ (2) = 126.23, $p < 0.0001$), as well as incentive ($\chi^2$ (1) = 7.2153, $p = 0.1073$). There was a significant interaction between signal degradation and condition ($\chi^2$ (2) = 30.375, $p < 0.0001$), which reflects the difference in performance in the threshold condition, where participants gave up in the incentive condition.

**Figure 4-1: Intelligibility (d-prime) scores for non-incentive (NI, dark colour) and incentive (I, light colour) conditions, with 2 added levels of signal distortion (+threshold=approx. 75% intelligibility, threshold=approx. 50% intelligibility)**




### 4.3.3 Frequency following response

Figure 4-2 shows correlation values for FFR. Stimulus degradation was found to increase the FFR, and this increase was larger in non-incentive conditions than in incentive conditions. In fact, the FFR for non-degraded conditions was essentially non-detectable. These findings are highly surprising and stand in contrast to those from Study 1 and previous research, which indicated that signal degradation is a primary factor resulting in a decrease, but not increase, in FFR. Furthermore, if the FFR is thought to show early attention effects, it would have been expected that an incentive would lead to a greater increase in tracking activity. Yet, the opposite was found; FFR in non-incentive conditions was enhanced compared to incentive conditions. Statistical analysis revealed a significant effect of signal degradation ($\chi^2$ (2) = 69.802, $p < 0.0001$) as well as incentive ($\chi^2$ (1) = 7.4962, p = 0.0061). The interaction was not significant ($\chi^2$ (2) = 5.7757, p = 0.0557).

**Figure 4-2: Signal reconstruction performance for FFR to target speaker for non-incentive (dark colour) and incentive (light colour) conditions, with 2 added levels of signal distortion (+threshold=approx. 75% intelligibility, threshold=approx. 50% intelligibility)**

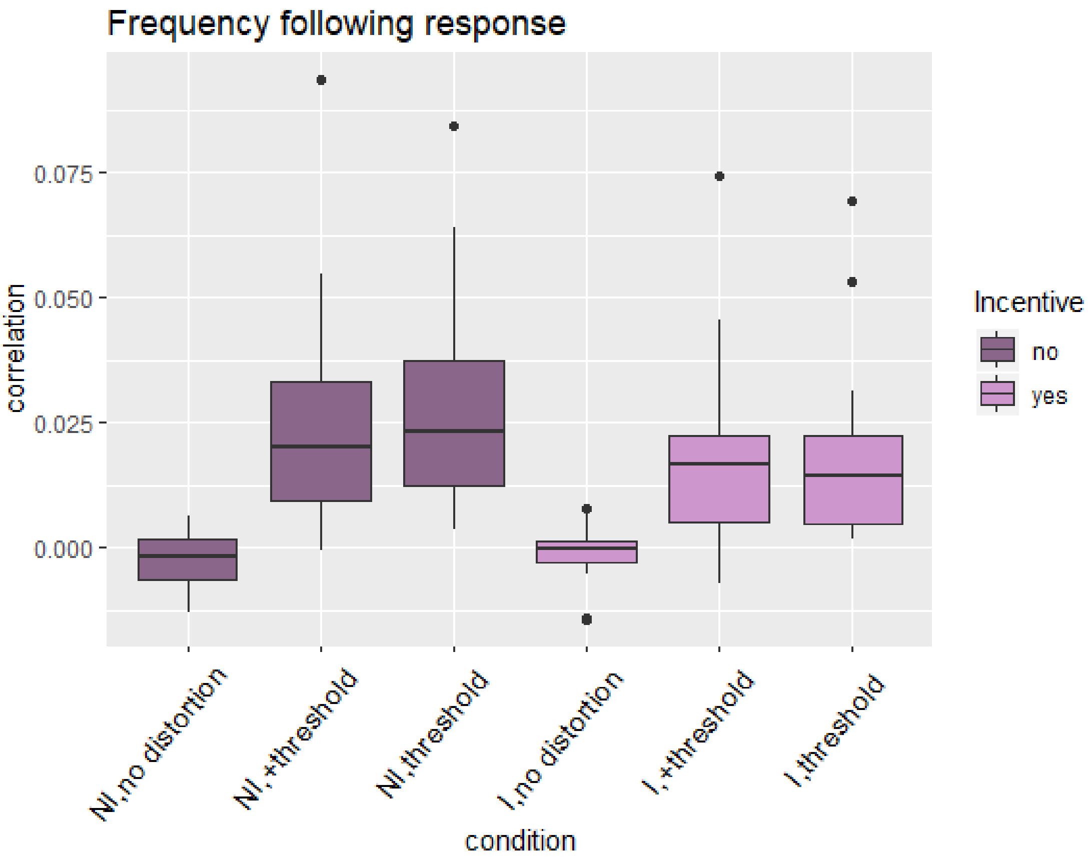
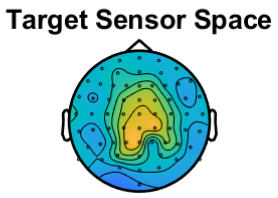
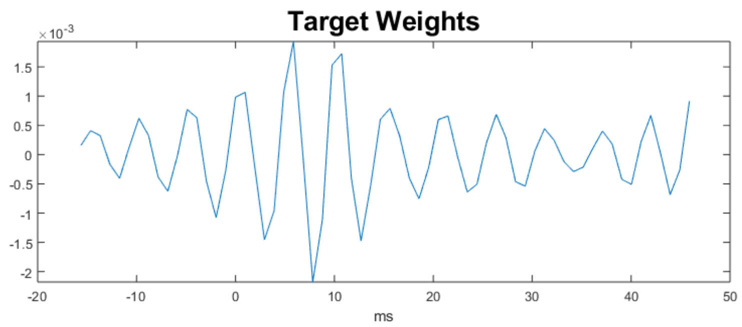


Figure 4-3 shows sensor space plots and target weights for FFR reconstruction (best electrode = Cz) within the time window of -15 to 45ms. Model weights peaked at approximately 8ms, which means that these time lags were the most important for reconstructing the stimulus. This indicates that the reconstructed waveform generated from the model was indeed generated at latencies typical of brainstem responses.

**Figure 4-3: Target and distractor sensor space plots (at maximum value) and mTRF model weights for FFR**




## 4.3.4 Cortical entrainment

Figure 4-4 shows cortical entrainment to target (red) and distractor speakers (blue), and Table 4-1 displays the mean coherence values. As in previous studies, entrainment to the target speaker was enhanced compared to entrainment to the distractor speakers. For both incentive and non-incentive conditions, entrainment to the target speaker was greatly increased with added degradation. Specifically, for non-incentive conditions entrainment in degraded conditions reached 174% (above threshold) and 196% (threshold) of the entrainment for non-degraded conditions. For incentive conditions, the same respective figures were 157% and 166%. Part of the reason for the smaller increase in entrainment for incentive conditions was the higher baseline (i.e. the non-degraded condition). While the presence of an incentive to perform better also seemed to increase cortical entrainment in non-degraded conditions, there was no such effect for conditions with added degradation. These results indicate that the addition of a degradation that preserves the speech envelope as well as periodicity, but flattens out the spectral detail, triggers enhanced cortical entrainment. This effect can be linked to similar observations in older, hearing-impaired and L2 listeners and might explain the increased entrainment in these groups. In contrast to results from Study 1, where greater degradation (noise and vocoding) resulted (if anything) in a drop in entrainment, these results indicate that even in normal-hearing subjects, degradation can increase entrainment. Potentially, the specific combination of a preserved amplitude envelope and fundamental frequency with decreased spectral detail triggers the cortical mechanism of increased entrainment. Statistical analysis on entrainment to the target speaker revealed a significant effect of signal degradation ($\chi^2$ (2) = 44.814, $p < 0.0001$), but not of incentive ($\chi^2$ (1) = 0, p = 0.9975). There was no significant interaction between signal degradation and incentive ($\chi^2$ (3) = 0.9374, p = 0.8164).

Compared to the entrainment for undistorted conditions measured in Study 1, levels recorded in the present study were low. The comparison between the two studies shows that average entrainment without incentive in this study only amounted to 59% of average entrainment in Study 1. Even in the corresponding condition with incentive, where entrainment was higher than without incentive, this figure only rises to 65%. One contributing factor for the higher entrainment in distortion-free conditions recorded in Study 1 could be the presence of only one (as opposed to two) distractor speaker. Arguably, the neural mechanism enhancing the target and suppressing the distractor needs to be stronger in a situation where the distractor is more salient than when the distractor consists of a blend of multiple voices. With added degradation in the present study, however, this trend of lower entrainment is strongly reversed: in threshold conditions, average entrainment amounted to 188% and 177% (without and with incentive, respectively) of entrainment found in the most difficult condition of Study 1 (which was aimed to represent threshold levels too). This again links to the contrasting effect of the degradation used in this experiment, which increased rather than decreased entrainment.

**Figure 4-4: Signal reconstruction performance for cortical entrainment across all conditions, for target and distractor speakers (top); boxplots and beeswarm plots to show variability of the data (bottom)**

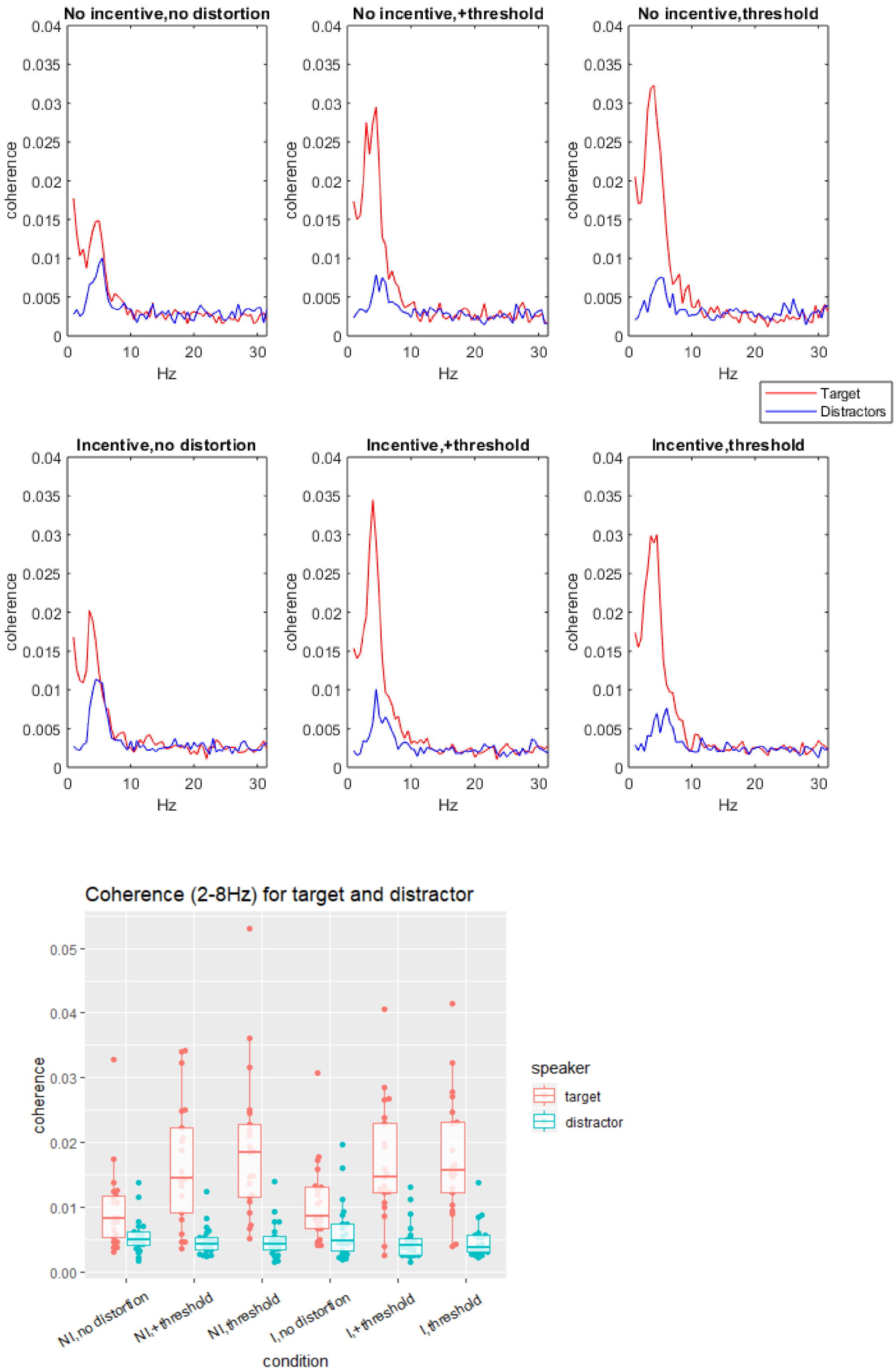


**Table 4-1: Mean coherence values to the target speaker per condition**

| **Mean coherence value (target) 2-8Hz** | |
|---|---|
| No incentive, no distortion | 0.0096 |
| No incentive, + threshold | 0.0167 |
| No incentive, threshold | 0.0189 |
| Incentive, no distortion | 0.0107 |
| Incentive, + threshold | 0.0168 |
| Incentive, threshold | 0.0177 |

Figure 4-5 shows target and distractor sensor space plots and model weights (best electrode = CP3) for target and distractor within the time window of 0 to 400ms. The target model weights showed a less clear but still visible P1-N1-P2-like effect. No such effect was observed for the distractor model weights. These results are similar to those from Study 1, indicating that the differences between active listening and ignoring a speaker are might be reflected in the model parameters.

**Figure 4-5: Target and distractor sensor space plots (at maximum value) and mTRF model weights for cortical entrainment**

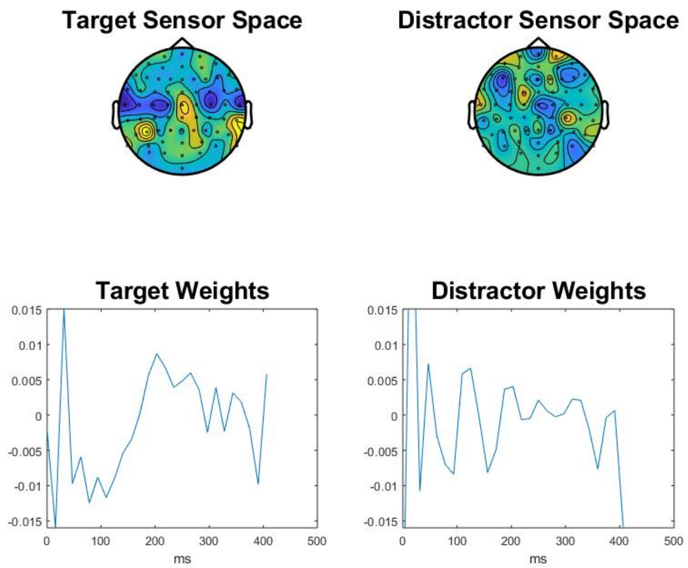
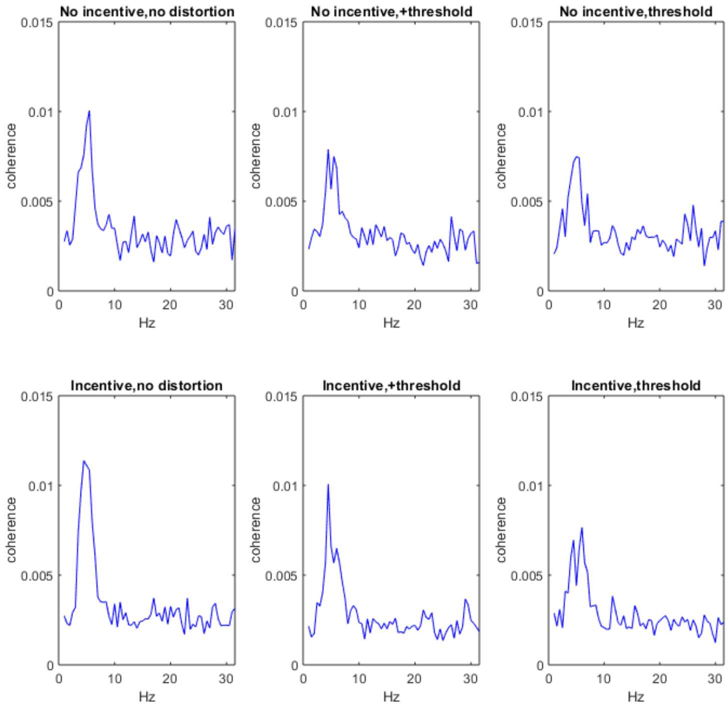


Figure 4-6 below shows coherence to the distractor speakers at an expanded scale. Entrainment to the distractors dropped with increasing distortion, in contrast to target speaker entrainment (which was enhanced in more severely degraded conditions). This trend was observed both for non-incentive and incentive conditions equally. These results indicate that increasing signal distortion not only resulted in an increase of target speaker entrainment, but also in the progressive suppression of the distractor speakers: the higher the amount of distortion in the signal, the greater the suppression of entrainment to the distractors. This mechanism could actively assist selective attention to the target by ameliorating distractor processing in adverse conditions. Statistical tests revealed a significant effect of degradation ($\chi^2$ (2) = 8.5685, p = 0.0138) but no significant effect of incentive ($\chi^2$ (1) = 0.3767, p = 0.5394). There was no significant interaction ($\chi^2$ (3) = 2.0758, p = 0.5568).

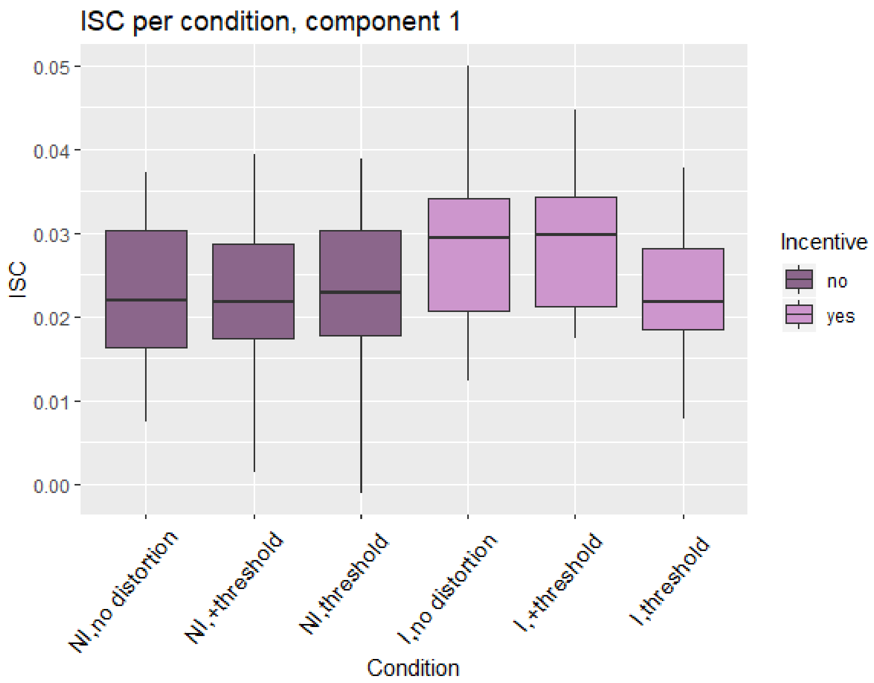
**Figure 4-6: Signal reconstruction performance for cortical entrainment across all conditions, distractor speakers only**



### 4.3.5 Intersubject correlation

Figure 4-7 shows intersubject correlation (ISC) by condition for the first component. While signal degradation did not affect ISC, higher ISC was observed in incentive conditions except for the threshold conditions, in which participants were behaviourally shown perform worse and give up due to the high level of difficulty. As ISC can be interpreted as a marker of attention, this is an indication for increased attention (and potentially effort) in conditions for which participants perceived to gain an advantage. The failure to observe increased ISC in the incentive threshold condition is in line with behavioural results and indicates that participants indeed gave up trying to reach a better score in order to finish the block early. Statistical analysis revealed a significant effect of incentive ($\chi^2$ (1) = 11.779, $p < 0.001$), but no effect of signal degradation ($\chi^2$ (2) = 3.5598, $p = 0.1687$). There was also an interaction between signal degradation and incentive ($\chi^2$ (4) = 10.641, $p = 0.031$). These results indicate that ISC was indeed significantly enhanced in incentive conditions. The interaction indicates that ISC as a marker of attention is dependent on the level of difficulty and degradation. If the degradation is perceived as too high (as in the threshold condition), no additional effort (and thus no additional attention) is expended.

**Figure 4-7: ISC per condition, component 1**




### 4.3.6 N400

In order to examine the distribution and latency of the N400 effect, non-parametric, cluster-based permutation tests were performed on the EEG data over a time window of 200ms to 600ms across all electrodes after the onset of the last word in order to test for significant clusters of time and sensor space (Maris & Oostenveld, 2007). Testing revealed a significant cluster over central regions for the difference between high and low cloze predictability sentences over latencies from 400 to 600ms ($p < 0.01$, alpha level of 0.05) (see Figure 4-8 below). There was also a significant interaction between condition (non-incentive vs. incentive) and sentence type (high vs. low cloze probability sentences) over a time window of 450ms to 600ms after word onset ($p < 0.05$, alpha level of 0.05) (see Figure 4-9 below).

**Figure 4-8: Permutation analysis of N400 effect for target speaker sentences with high and low cloze predictability, for latencies between 200ms and 600ms**

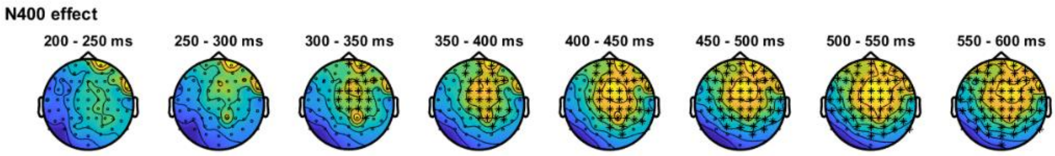
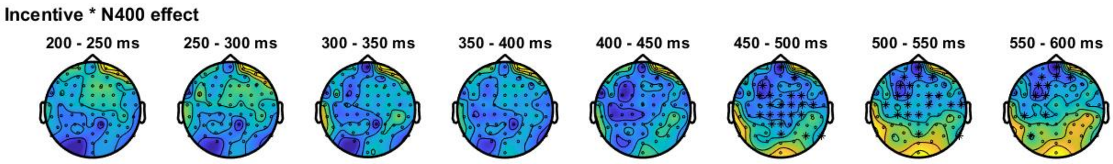
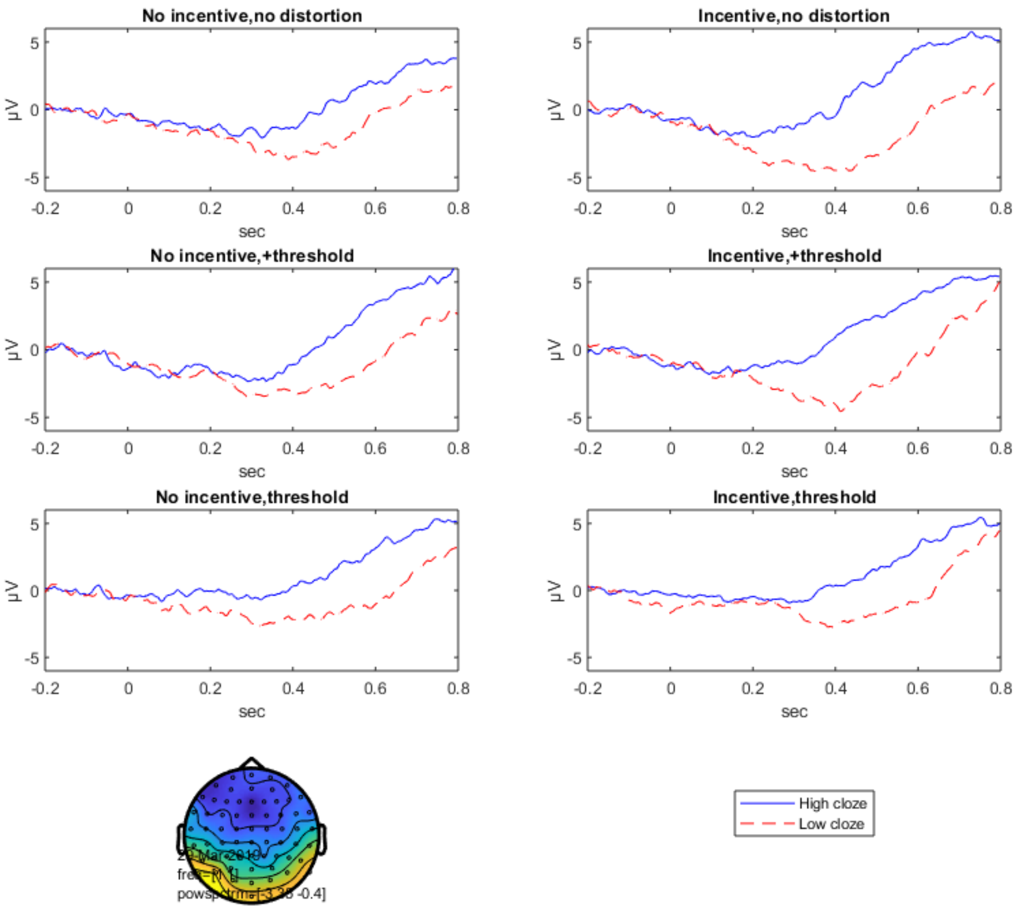


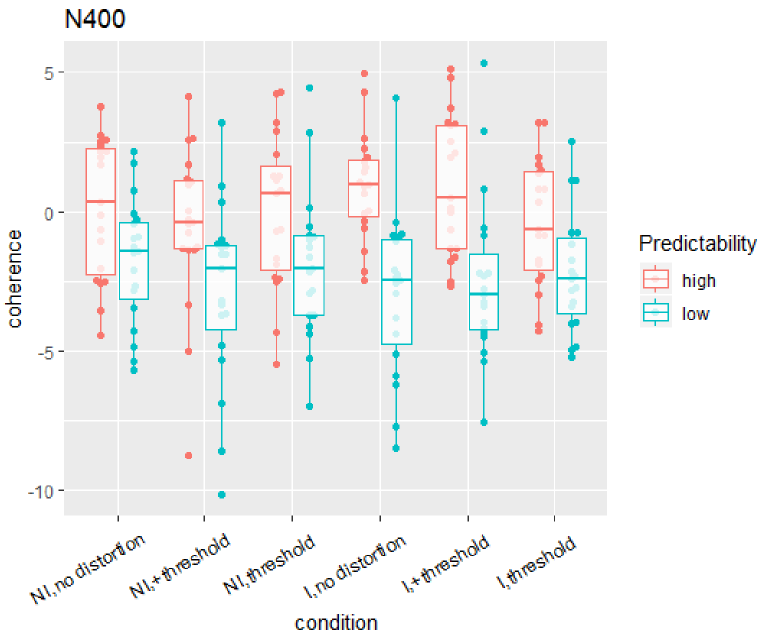

**Figure 4-9: Permutation analysis of interaction between N400 effect and incentive**




The main statistical analysis was performed using a more narrow time window (300-500ms) and smaller midline electrode set (Fz, FCz, Cz, CPz, and Pz). Figure 4-10 shows voltage over time after the onset of the last word for high cloze probability (blue) and low cloze probability (red) sentences (mean results at FCz). For all conditions, an N400 effect was observed (i.e. different levels of processing for high vs. low cloze probability sentences around 400ms). However, signal degradation did not affect the N400 effect, with the most degraded conditions showing the same effect as the non-degraded conditions. Visually, incentive did increase the N400 effect, as there was slightly increased activity for low-cloze probability sentences in incentive conditions which would indicate greater processing effort with incentive. Statistical analysis over a time window from 300 to 500ms revealed that sentence type (high vs. low cloze probability) was significant ($\chi^2$ (1) = 41.846, p < 0.0001). However, neither incentive ($\chi^2$ (1) = 0.0932, p = 0.7601) nor signal degradation showed any significant effects ($\chi^2$ (2) = 0.5439, p = 0.7619). There was no significant interaction between the factors. This means that, despite the visual differences in the N400 effect between non-incentive and incentive conditions, this effect was not strong enough to be statistically significant.

**Figure 4-10: N400 effect for target speaker sentences with high and low cloze predictability sentences, across all conditions (top); boxplots and beeswarm plots to show variability of the data (bottom)**






## 4.4 Discussion

This study investigated the effects of incentive as well as a particular type of signal degradation on various neurophysiological measures. The observed effect of incentive was overall not very large and interacted with degradation, but was present across a range of different behavioural and neurophysiological measures. The degradation used in this study also impacted various measures, and most importantly resulted in a substantial increase in cortical entrainment in normal-hearing subjects despite reduced intelligibility.

The behavioural results from this study serve as a first indicator of the impact of signal degradation and incentive on speech perception performance. Results across both incentive and non-incentive conditions showed a similar trend: performance overall dropped with increasing degradation, and was lowest in conditions where degradation was at the individual threshold level of 50% speech perception as expected. However, while generally there were no differences between incentive and non-incentive conditions, the most difficult incentive conditions stood out as subjects actually performed worse than without an incentive. While this result at first glance seems counterintuitive, it can be explained with the propositions laid out by Motivational Intensity Theory (Brehm & Self, 1989; Richter, 2013; Wright, 2008). This theory emphasises the importance not only of task difficulty, but also the principle of energy conservation for energy investment. When a task is subjectively perceived as too difficult, the effort invested is low, as success is perceived as impossible (Wright & Franklin, 2004). While subjects in our experiment had nothing to gain – but also nothing to lose – in the non-incentive condition, they might have judged the threshold conditions as too difficult and not worth the effort when confronted with the requirement to collect points. Thus, whereas they kept trying (and therefore achieved scores even similar to easier conditions) in the non-incentive difficult condition, participants gave up entirely in incentive conditions when they assessed the score required to finish ahead of time to be impossible to reach.

Results from the analyses of cortical coherence to the stimulus indicate that higher levels of degradation increased entrainment substantially. While previous research, and Study 1 presented in this thesis, have suggested that the relationship between intelligibility and entrainment is generally positive (i.e. better intelligible stimuli evoke higher entrainment) (Ding et al., 2014; Peelle, Gross, & Davis, 2013; Vanthornhout et al., 2018), it is becoming increasingly clear that this relationship is neither linear nor all-encompassing. Studies with ageing, hearing-impaired and non-native populations in particular have found enhanced entrainment in these groups compared to normal-hearing young adults, despite typically worse (or at best equal) performance (Goossens et al., 2018; Millman et al., 2017; Presacco et al., 2016a, 2016b; Song & Iverson, 2018). Thus, there is a growing body of evidence to suggest that higher entrainment is not always related to higher attention or better speech perception performance.

The differences in entrainment observed between Study 1 and the present study suggest that the degradation's characteristics critically influence not only the magnitude of entrainment, but also whether entrainment under degraded conditions is increased or decreased in relation to a degradation-free condition. In Study 1, the degradation type used was speech-shaped noise. The present study, in contrast, used a degradation that, while retaining the f0 and amplitude envelope of the target signal, flattened out the details of the target spectrum. Previous research has demonstrated that spectral details are important in particular for speech perception in adverse conditions. Spectral smearing (reduced frequency resolution) increases speech reception thresholds progressively in normal-hearing people (Keurs, Festen, & Plomp, 1989), and makes speech presented in a fluctuating masker less intelligible (Bernstein & Brungart, 2011). The present study suggests that these perceptual effects are also reflected in neural processing by the auditory cortex: reducing the spectral detail triggers increased reliance on envelope tracking despite worse speech perception performance.

The results of higher entrainment in conjunction with reduced processing of spectral detail can provide an explanation for findings with the elderly, hearing-impaired and L2 listeners (e.g. Presacco et al., 2016b; Song & Iverson, 2018). These groups have also demonstrated increased levels of entrainment despite worse speech intelligibility compared to normal-hearing, young L1 listeners. The subjective reports that these listener groups need to 'work harder' to recognize speech in difficult conditions (Bernarding, Strauss, Hannemann, Seidler, & Corona-Strauss, 2013; Ohlenforst et al., 2017; Tun, McCoy, & Wingfield, 2009), as well as the direct link between reduced frequency resolution and lower speech perception in hearing-impaired people (Glasberg & Moore, 1986; Summers & Leek, 1994) suggest that the mechanism observed might be the same as measured in the present study with normal-hearing people. These results suggest that the unavailability and/or processing of spectral detail might be a decisive factor leading to higher entrainment without attentional or intelligibility benefits.

An interesting finding was the countertrend observed for distractor speaker coherence: while target speaker entrainment was increased with added degradation, distractor speaker entrainment fell. Although most of the literature on cortical entrainment has focused on the target speaker, some studies have investigated the neural processing of distractors as well (Olguin, Bekinschtein, & Bozic, 2018; Woestmann, Lim, & Obleser, 2017). One such recent study has found that under adverse listening conditions (i.e. in negative SNRs), neural distractor speaker representation is actively suppressed in order to counteract effects of bottom-up attention which would favour the distractor speaker (Fiedler et al., 2019). Findings from the present study expand these results with the insight that distractor speaker entrainment suppression is adaptive and progressively strengthened with increasing adversity.

Of all measures extracted from the EEG signal, intersubject correlation (ISC) was the one closest to a direct measure of effort as proposed in the literature. As previous EEG and MEG studies have shown, subjects' electroencephalographic responses to stimuli are more similar to one another in situations where they are more engaged, or paying more attention (Cohen & Parra, 2016; Dmochowski et al., 2012; Ki et al., 2016). Thus, ISC can serve as an indicator of auditory attention. Since effort can be defined as the "deliberate allocation of resources" (Pichora-Fuller et al., 2016), higher attention can be the result of increased effort put into a task. While increased attention might not necessarily be evoked by higher effort but also by other factors such as different stimulus materials or different listening environments (e.g. more engaging, more interesting), the present experiment sought to eliminate such other factors by holding everything else equal across incentive and non-incentive conditions. Thus, differences measured between conditions with and without incentive ought to be attributable to this specific experimental manipulation. Against this background, the significant effect of incentive on ISC confirms the original hypothesis. These results indicate that subjects did indeed pay more attention, and put more effort into the task when there was an incentive to perform better.

In contrast to ISC, no significant effect of incentive was observed for cortical entrainment. One of the potential reasons for the absence of this effect could be that the incentive effect was too small in comparison to the large effect of signal degradation. Another possibility would be that this dimension of attention and effort, which appears enhanced for incentive conditions, is not reflected in cortical entrainment. Support for this explanation comes from Study 2 findings, which also failed to record differences between conditions with different hearing aid algorithms that subjectively and behaviourally aided speech perception in adverse conditions.

A puzzling result from this study was entrainment to the frequency following response (FFR). In various previous studies it has been found that the FFR is highly sensitive to degradation (e.g. Cunningham et al., 2001; Russo et al., 2004) – which was partly the reason that used to necessitate many repetitions of the same stimulus to allow for the elimination of noise in the signal. Results from Study 1 affirmed these findings: while FFR tracking to the distractor speaker (which was never degraded) remained at a high level throughout all conditions, entrainment to the target speaker was significantly decreased in conditions with added noise and even eliminated in the vocoded condition (where the f0 was erased from the signal). Yet, in the present study FFR tracking was only measured in conditions with degradation, but not in degradation-free ones. Based on results from Study 1, the opposite would be expected. A possible reason for this result might be the use of the same voice as distractor speakers: potentially, entrainment was 'washed out' by the identical voices in the easiest conditions. Yet, these results still remain difficult to reconcile with findings from Study 1 and previous research. Future investigations could explore the effect of distractors (i.e. same sex vs. opposite sex, same vs. different voice) on FFR tracking. Another possible explanation for the lack of FFR tracking in non-degraded conditions is the use of the particular degradation type in the present study. Potentially, the combination of a flattened modulation spectrum and the preservation of the fundamental frequency might enhance FFR tracking. However, in this case it would still remain unclear why no FFR tracking was obtained in non-degraded conditions.

N400 results did not indicate significant effects of either signal degradation or incentive. These findings, similarly to entrainment, differ from that of Study 1, where a decreased N400 effect was observed in conditions with greater distortion. A visual inspection, supported by results from permutation statistics, suggested that incentive increased the N400 effect, which would indicate higher levels of processing effort for lexical integration (Chwilla et al., 1995). However, statistical analysis did not support these hypotheses beyond the clear significance of word predictability (i.e. the different processing of high vs. low cloze probability sentences).

In sum, findings from this study provide evidence to explain the increased cortical entrainment in hearing-impaired, older and L2 listeners measured in previous research and Study 2. It was found that cortical entrainment in normal-hearing subjects can be increased in degraded conditions with reduced intelligibility as well. Crucially, it might be the reduced ability to process spectral detail despite the presence of periodicity and the amplitude envelope that triggers the cortical mechanism of increased entrainment. In contrast to signal degradation, incentive was not found to have broad impact on neurophysiological measures beyond intersubject correlation. This might be due to either its small effect size, or the fact that a measure such as cortical entrainment does not capture these specific aspects of attention and processing effort.

# Chapter 5: General discussion

Selective auditory attention has been a topic of research for decades. Nevertheless, the neural mechanisms underpinning the 'cocktail party problem' of selective attention to a speaker in a multi-talker environment are still poorly understood. Previous behavioural and subjective evidence has shown that, while normal-hearing people have listening abilities which are surprisingly robust to degradation and noise, hearing-impaired people and speech-recognition machines are much more vulnerable in adverse conditions. Therefore, a better understanding of the ways in which people selectively attend to speech in challenging conditions would benefit not only the research community, but could also inform and inspire the development of better hearing aids and non-human speech recognizers. Recently, analysis techniques based on machine learning have made it possible to reliably decode auditory attention from the neural signal. The research conducted for the present thesis has built on this line of research and used a toolbox of neurophysiological measures to address questions of attention and listening effort under various circumstances and in different populations.

Study 1 was used to develop a battery of measures to investigate auditory attention from a multidimensional, multi-stage processing perspective, as well as to apply this battery of measures in an experiment with different types and degrees of masking. In addition to gathering a behavioural measure of speech perception, the EEG signal was analysed to extract measures corresponding to different stages of neural processing, from early responses in the brainstem through cortical entrainment to late stages of semantic and lexical processing (N400). Besides the development of the methodological framework, this study yielded results indicative of different factors influencing the measures. While very early stages of neural speech processing were strongly tied to 'bottom-up' factors such as stimulus characteristics (such as degradation), late stages were more closely associated with 'top-down' factors such as intelligibility.

Study 2 examined differences in neural speech processing and attention between normal-hearing and hearing-impaired people. Additionally, this study sought to investigate the effects of algorithms commonly implemented in hearing aids on neurophysiological measures of speech processing. Findings from this study pointed to the intricate relationship between age, hearing impairment and neural entrainment. While intersubject correlation, a neurophysiological measure of attention, did not differ significantly between the groups, cortical entrainment to the speech envelope was enhanced in the hearing-impaired group despite similar or worse levels of behavioural speech recognition performance. In contrast to young normal-hearing adults, for whom greater entrainment is generally associated with higher attention and intelligibility, older hearing-impaired people displayed far greater cortical entrainment despite equal or lower behavioural performance. These results suggest that age and hearing impairment may alter the mechanisms of selective auditory attention, potentially through less impaired frequency selectivity and the resulting greater reliance on envelope cues, and/or less effective suppression of distractor sounds.

Study 3 investigated the impact of motivation as a top-down factor, as well as the effect of a specific type of stimulus degradation as a bottom-up factor of influence on neurophysiological measures. The degradation used in this study triggered higher entrainment in degraded conditions in normal-hearing subjects despite worse speech recognition performance, similarly to effects observed in older, hearing-impaired subjects in Study 2. These results suggest that specifically the reduced ability to process spectral detail despite the availability of f0 and amplitude envelope information might be responsible for the cortical mechanism of higher entrainment in degraded conditions. While the manipulation of an incentive to perform better in this study impacted several behavioural as well as neurophysiological measures, this effect was smaller and generally interacted with degradation.

Methodologically, research on auditory attention is commonly rooted within one of two fields: (bio-)engineering and psychology. The aim of this thesis was to bring methodologies and approaches from both fields together and investigate concepts such as effort and attention with a multi-lens perspective. One advantage of the most recent, machine-learning based EEG analysis techniques of attention decoding is that they do not require the use of carefully constructed sentences. Particularly in the study with the hearing-impaired group, this feature was exploited to create a more realistic listening experiment under laboratory conditions and demonstrate the viability of real-life stimuli taken from a radio program for successfully conducting EEG research.

The central measure investigated in all studies of the current thesis was cortical entrainment. In line with previous research, it was demonstrated that the synchronization in low-frequency regions (2-8Hz) between the speech envelope and the neural signal shows an attention advantage, i.e. that cortical entrainment was consistently enhanced to a target speaker compared to a distractor speaker. In young normal-hearing subjects, higher cortical entrainment in general appeared to be a positive, ‘desirable’ marker, as it was linked to higher speech intelligibility. However, age and hearing impairment appeared to reverse some of the ‘more is better’ principles of cortical entrainment. In this (and previous) studies, older, hearing-impaired people displayed greatly increased cortical entrainment levels of up to 200%, compared to a young, normal-hearing group. At the same time, other neurophysiological measures of attention as well as behavioural performance did not indicate that the hearing-impaired group gained any advantage in terms of higher speech recognition or increased attention from this vastly enhanced entrainment. Findings from Study 3 indicate that an explanation for this higher entrainment could be the decreased ability to process spectral and phonetic details of the speech, and a resulting greater reliance on the speech envelope cues. Other factors at play might be altered cognitive pathways or a reduced ability to filter out distractor sounds in an ageing and/or hearing-impaired brain.

Particularly these findings from Study 2 and Study 3 point to interesting directions for future research. While it was demonstrated here that the combination of age and hearing impairment increases neural entrainment, it would be desirable to disentangle the effect of these two factors from one another. While some studies have already started to investigate neural entrainment in these groups (Goossens et al., 2018; Millman et al., 2017; Presacco et al., 2016a, 2016b), more research is needed comparing older, normal-hearing with older, hearing-impaired as well as young, normal-hearing with young, hearing-impaired groups. Such findings could advance our understanding of the cognitive and neural changes that occur with ageing and hearing impairment separately. Another interesting path for investigation in these specific populations is distractor entrainment. In Study 2 of this thesis, the distractor used was a multi-talker babble noise that primarily provided energetic, but no informational masking. Based on these findings and those from previous research (Petersen et al., 2017), experiments with hearing-impaired people could be conducted in dichotic listening situations (i.e. intelligible distractor speakers) so as to investigate whether the processing of the distractor appears enhanced in similar ways as that of the target. This might indicate that some of the difficulties hearing-impaired people experience in noisy situations could have their roots in a weakened neural distractor speaker suppression mechanism.

This thesis has, for the first time, demonstrated that cortical entrainment can be increased in degraded conditions, even in normal-hearing subjects. Together with findings from the hearing-impaired group, as well as previous research, these results might be indicative of a processing balance of different speech cues that is disturbed when certain cues (in this case spectral detail) are unavailable. Perhaps as a compensatory mechanism, cortical tracking of the broadband envelope is increased under such conditions. Rather than aiding with speech comprehension, though, this increased tracking activity is associated with worse speech perception. These findings are particularly important as they point to a specific mechanism by which increased cortical entrainment, hearing impairment and worse speech perception are related to one another. They further show that increased entrainment in degraded conditions does not, in itself, characterize hearing impairment, but can be elicited in normal-hearing people as well.

In addition to these significant findings related to particular measures, the strength and novelty of the work conducted for this thesis lies in its approach: while most previous research, particularly relating to listening effort and auditory attention, has concentrated on operationalizing these complex concepts through one measure, this thesis demonstrates that the combination of different measures within a single experiment provides highly valuable insights. For example, Study 2 showed that hearing aid algorithms such as noise reduction and directional microphones reduce subjective listening effort as well as increase speech comprehension performance, in both hearing-impaired as well as normal-hearing people. Neurophysiological measures, however, were not significantly affected by the type of algorithm activated. Furthermore, while there were large group differences between the normal-hearing and the hearing-impaired for cortical entrainment, intersubject correlation as a global measure of attention showed no such group differences. These differentiated findings for different measures highlight the multi-dimensional, multi-stage nature of concepts such as listening effort and attention. Future research can build on the toolbox of measures developed in this thesis, and potentially expand the framework to include multi-modal measures such as the combined recording of EEG and pupillometry. Such studies, building on the idea of multi-dimensional investigation developed in this thesis, could shed light on the different aspects of auditory attention measured by different methodologies, and further refine our understanding of speech processing in the brain.

Beyond the contribution to academic research, the work conducted for this thesis has impacted industrial research and product development, as well as contributed to public understanding of and engagement with scientific research. Specifically, the project for Study 2 was carried out in collaboration with a leading hearing care providers. The EEG research conducted for this study introduced the company, for the first time, to neurophysiological measures and their applicability as well as feasibility to assess various features implemented in hearing devices. After successful completion of the project, the company purchased equipment for similar neurophysiological measurements and now continues to use it for research into and evaluation of hearing aids in conjunction with more traditional behavioural measures of speech comprehension. Additionally, this work has been presented at various public and institutional events, raising awareness for research on auditory attention, neural language processing and hearing loss.

In sum, the current thesis investigated the relationship between signal degradation, auditory attention and effort in different populations, using a toolbox of neurophysiological, behavioural and subjective measures. Specifically, different types and levels of degradation impact online measures of neural speech processing in diverging ways. While early measures are more affected by peripheral, bottom-up factors, late measures are more closely associated with top-down cognitive and central processes. In older, hearing-impaired people, neural markers of speech processing (specifically cortical entrainment) are affected in opposite ways as would be predicted from data on young, normal-hearing people. Findings from this thesis suggest that this might be due to the reduced ability to process spectral details of the speech. While bottom-up factors such as type and degree of signal degradation play a key role in shaping the neural response, top-down factors such as motivation and effort also influence online markers of speech processing. Especially the differences found between normal-hearing and hearing-impaired people as well as the extent of distractor speech processing are interesting lines of investigation for future research on auditory attention.

# References


Action on Hearing Loss. (2018). Facts and figures.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Alain, C., McDonald, K., & Van Roon, P. (2012). Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. *Hearing Research*, *283*(1–2), 126–135. https://doi.org/10.1016/j.heares.2011.10.007

Alcántara, J. I., Moore, B. C. J., Kühnel, V., & Launer, S. (2003). Evaluation of the noise reduction system in a commercial digital hearing aid. *International Journal of Audiology*, *42*(1), 34–42. https://doi.org/10.3109/14992020309056083

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2013). Auditory Brainstem Response to Complex Sounds Predicts Self-Reported Speech. *Journal of Speech, Language, and Hearing Research*, *56*(February 2013), 31–43. https://doi.org/10.1044/1092-4388(2012/12-0043)in

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural Timing Is

Linked to Speech Perception in Noise. *Journal of Neuroscience*, *30*(14), 4922–4926. https://doi.org/10.1523/JNEUROSCI.0107-10.2010

Arbogast, T. L., Mason, C. R., & Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2169–2180. https://doi.org/10.1121/1.1861598

Arlinger, S., & Gustafsson, H.-A. (1991). Masking of speech by amplitude-modulated noise. *Journal of Sound and Vibration*, *151*(3), 441–445.

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–803. https://doi.org/10.1038/nn.2810

Ashmore, J. (2008). Cochlear Outer Hair Cell Motility. *Physiol. Rev.*, *88*, 173–210. https://doi.org/10.1152/physrev.00044.2006.

Baltzell, L. S., Horton, C., Shen, Y., Richards, V. M., D’Zmura, M., & Srinivasan, R. (2016). Attention selectively modulates cortical entrainment in different regions of the speech spectrum. *Brain Research*, *1644*, 203–212. https://doi.org/10.1016/j.brainres.2016.05.029

Barker, F., Mackenzie, E., Elliott, L., Jones, S., & de Lusignan, S. (2016). Interventions to improve hearing aid use in adult auditory rehabilitation. *The Cochrane Database of Systematic Reviews*, *8*.

https://doi.org/10.1002/14651858.CD010342.pub2

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America, 139*(3), EL51–EL56. https://doi.org/10.1121/1.4942628

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bentler, R. A. (2005). Effectiveness of Directional Microphones and Noise Reduction Schemes in Hearing Aids: A Systematic Review of the Evidence. *Journal of the American Academy of Audiology*, *16*(7), 473–484. https://doi.org/10.3766/jaaa.16.7.7

Bentler, R., & Chiou, L.-K. (2006). Digital Noise Reduction: An Overview. *Trends in Amplification*, *10*(2), 67–82. https://doi.org/10.1177/1084713806289514

Bentler, R., Wu, Y. H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *International Journal of Audiology*, *47*(8), 447–460. https://doi.org/10.1080/14992020802033091

Bernarding, C., Strauss, D. J., Hannemann, R., & Corona-Strauss, F. I. (2012). Quantification of listening effort correlates in the oscillatory EEG activity: A feasibility study. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 4615–4618. https://doi.org/10.1109/EMBC.2012.6346995

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: Influence of age and hearing impairment. *Brain Research Bulletin*, *91*, 21–30. https://doi.org/10.1016/j.brainresbull.2012.11.005

Bernstein, J. G. W., & Brungart, D. S. (2011). Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. *The Journal of the Acoustical Society of America*, *130*(1), 473–488. https://doi.org/10.1121/1.3589440

Bernstein, J. G. W., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *120*, 3929–3945.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, *77*(1), 1–10. https://doi.org/10.1016/j.bandc.2011.07.006

Bonte, M., Parviainen, T., Hytönen, K., & Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex*, *16*(1), 115–123. https://doi.org/10.1093/cercor/bhi091

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in*

*Neuroscience*, *12*(MAR), 1–13. https://doi.org/10.3389/fnins.2018.00152

Boymans, M., & Dreschler, W. A. (2000). Field Trials Using a Digital Hearing Aid with Active Noise Reduction and Dual-Microphone Directionality: Estudios de campo utilizando un audifono digital con reduccion activa del ruido y micrófono de direccionalidad dual. *International Journal of Audiology*, *39*(5), 260–268. https://doi.org/10.3109/00206090009073090

Brehm, J. W., & Self, E. A. (1989). The Intensity Of Motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.psych.40.1.109

British Society of Audiology. (2017). *Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking*.

Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current Biology*, *28*(5), 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication*, *52*(11–12), 980–995. https://doi.org/10.1016/j.specom.2010.08.010

Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *The Journal of the Acoustical Society of America*, *92*(6), 3132–3139. https://doi.org/10.1121/1.404209

Bronkhorst, Adelbert W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception, and Psychophysics*, *77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Brons, I., Houben, R., & Dreschler, W. A. (2015). Acoustical and Perceptual Comparison of Noise Reduction and Compression in Hearing Aids. *Journal of Speech, Language, and Hearing Research*, *58*, 1363–1376. https://doi.org/10.1044/2015_JSLHR-H-14-0347

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(1001), 1101–1109. https://doi.org/10.1007/0-387-22794-6_17

Buchholz, J. M., Weissner, A., Beechey, T., Galloway, J., Keidser, G., & Freeston, K. (2016). Estimating realistic speech levels for virtual acoustic environments. *Journal of the Acoustical Society of America*, *140*(4), 3437–3437.

Burkard, R. F., & Sims, D. (2002). A Comparison of the Effects of Broadband Masking Noise on the Auditory Brainstem Response in Young and Older Adults. *American Journal of Audiology*, *11*(1), 13–22.

https://doi.org/10.1044/1059-0889(2002/004)

Campbell, K. L., Shafto, M. A., Wright, P., Tsvetanov, K. A., Geerligs, L., Cusack, R., … Villis, L. (2015). Idiosyncratic responding during movie-watching predicted by age differences in attentional control. *Neurobiology of Aging*, *36*(11), 3045–3055. https://doi.org/10.1016/j.neurobiolaging.2015.07.028

Cantlon, J. F., & Li, R. (2013). Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. *PLoS Biology*, *11*(1). https://doi.org/10.1371/journal.pbio.1001462

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-Dependent Encoding in the Human Auditory Brainstem Relates to Hearing Speech in Noise: Implications for Developmental Dyslexia. *Neuron*, *64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009.10.006

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907373

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274–285.

Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *ENeuro*, *3*(December), 1–11. https://doi.org/10.1523/ENEURO.0203-16.2016

Cole, L. w. (1911). The relation of strength of stimulus to rate of learning in the chick. *Journal of Animal Behavior*, *1*(2), 111–124.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Corbett, M. (2015). From law to folklore: Work stress and the Yerkes-Dodson Law. *Journal of Managerial Psychology*, *30*(6), 741–752. https://doi.org/10.1108/JMP-03-2013-0085

Cord, M. T., Surr, R. K., Walden, B. E., & Olson, L. (2002). Performance of directional microphone hearing aids in everyday life. *J Am Acad Audiol*, *13*(6), 295–307.

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience*, *35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Crosse, Michael J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*(November), 1–14. https://doi.org/10.3389/fnhum.2016.00604

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001).

Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology*, *112*(5), 758–767. https://doi.org/10.1016/S1388-2457(01)00465-5

Dai, L., Best, V., & Shinn-Cunningham, B. G. (2018). Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. *Proceedings of the National Academy of Sciences*, *115*(14), 201721226. https://doi.org/10.1073/pnas.1721226115

Danermark, B., & Gellerstedt, L. C. (2004). Psychosocial work environment, hearing impairment and health. *International Journal of Audiology*, *43*(7), 383–389. https://doi.org/10.1080/14992020400050049

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America*, *114*(5), 2913–2922. https://doi.org/10.1121/1.1616924

Das, N., Van Eyndhoven, S., Francart, T., & Bertrand, A. (2016). Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2016-Octob*(637424), 77–80. https://doi.org/10.1109/EMBC.2016.7590644

Davis, H., & Zerlin, S. (1966). Acoustic relations of the human vertex potential. *The*

*Journal of Neuroscience*, *39*(109), 109–116. https://doi.org/10.1121/1.1909858

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear & Hearing*, *35*(6), 600–610.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Research*, *70*, 80–90.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., … Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, *88*, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. https://doi.org/10.1073/pnas.1205381109

Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *Journal of Neuroscience, 33*(13), 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience, 8*, 1–7. https://doi.org/10.3389/fnhum.2014.00311

Divenyi, P. (2005). Humans glimpse, too, not only machines (hommage a Martin Cooke). *Forum Acusticum*, (June), 1533–1538. Retrieved from http://webistem.com/acoustics2008/acoustics2008/cd1/data/fa2005-budapest/paper/543-0.pdf

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, *5*, 1–9. https://doi.org/10.1038/ncomms5567

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? *Frontiers in Human Neuroscience*, *6*(May), 1–9. https://doi.org/10.3389/fnhum.2012.00112

Dodson, J. D. (1915). The relation of strength of stimulus to rapidity of habit-formation in the kitten. *Journal of Animal Behavior*, *5*(4), 330–336.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behav. Brain Sci.*, *11*, 357–427.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, *76*(1), 87–96. https://doi.org/10.1121/1.391011

Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *The Journal of the Acoustical Society of America*, *91*, 2110–2121. https://doi.org/10.1121/1.403697

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, *74*(3), 739–743. https://doi.org/10.1121/1.389859

Edwards, B. (2007). The Future of Hearing Aid Technology. *Trends in Amplification*, *11*(1), 31–45. https://doi.org/10.1177/1084713806298004

Ernst, S. M. A., & Moore, B. C. J. (2012). The role of time and place cues in the detection of frequency modulation by hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *131*, 4722–4731. https://doi.org/10.1121/1.3699233

Etard, O., Kegler, M., Braiman, C., Forte, A., & Reichenbach, T. (2019). Decoding of selective attention to continuous speech from the human auditory brainstem response. *NeuroImage*, *200*(May), 1–11. https://doi.org/10.1016/j.neuroimage.2019.06.029

Faulkner, A., Rosen, S., & Wilkinson, L. (2001). Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear and Hearing*, *22*(5), 431–438. https://doi.org/10.1097/00003446-200110000-00007

Favre-Felix, A., Graversen, C., Dau, T., & Lunner, T. (2017). Real-time estimation of eye gaze by in-ear electrodes. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 4086–4089. https://doi.org/10.1109/EMBC.2017.8037754

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Fiedler, L., Woestmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, *186*, 33–42. https://doi.org/10.1016/j.neuroimage.2018.10.057

Forte, A. E., Etard, O., & Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *ELife*, *6*, 1–12. https://doi.org/10.7554/eLife.27203

Fredelake, S., Holube, I., Schlueter, A., & Hansen, M. (2012). Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *International Journal of Audiology*, *51*(4), 299–308. https://doi.org/10.3109/14992027.2011.645075

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, *19*.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–3588. https://doi.org/10.1121/1.428211

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, *25*, 355–373.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. https://doi.org/10.1121/1.1381538

Fuellgrabe, C. (2013). Age-Dependent Changes in Temporal-Fine- Structure

Processing in the Absence of Peripheral Hearing Loss. *American Journal of Audiology*, *22*(December), 313–316. https://doi.org/10.1044/1059-0889(2013/12-0070)American

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, *156*(April), 435–444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Gagné, J. P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, *21*, 1–25. https://doi.org/10.1177/2331216516687287

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport*. https://doi.org/10.1097/00001756-199511270-00021

Galbraith, G. C., & Arroyo, C. (1993). Selective attention responses and brainstem. *Biological Psychology*, *37*, 3–22.

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., & Mullen, T. A. (1998). Brain stem frequency-following response to dichotic vowels during attention. *NeuroReport*, *9*(8), 1889–1893. https://doi.org/10.1097/00001756-199806010-00041

Galbraith, G. C., & Doan, B. Q. (1995). Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *International Journal of Psychophysiology*, *19*(3), 203–214. https://doi.org/10.1016/0167-8760(95)00008-G

Gendolla, G. H. E., & Wright, R. A. (2005). Motivation in social settings: Studies of Effort-Related cardiovascular arousal. In *Social Motivation: Conscious and Unconscious Processes* (pp. 71–90). https://doi.org/10.1017/CBO9780511735066.007

Gendolla, G. H. E., Wright, R. A., & Richter, M. (2012). Effort Intensity: Some Insights From the Cardiovascular System. In *The Oxford Handbook of Human Motivation*. https://doi.org/10.1093/oxfordhb/9780195399820.013.0024

Getzmann, S., Golob, E. J., & Wascher, E. (2016). Focused and divided attention in a simulated cocktail-party situation: ERP evidence from younger and older adults. *Neurobiology of Aging, 41*, 138–149. https://doi.org/10.1016/j.neurobiolaging.2016.02.018

Glasberg, B. R., & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, *79*(4), 1020–1033. https://doi.org/10.1121/1.393374

Golestani, N., Hervais-Adelman, A., Obleser, J., & Scott, S. K. (2013). Semantic versus perceptual interactions in neural processing of speech-in-noise. *NeuroImage, 79*, 52–61. https://doi.org/10.1016/j.neuroimage.2013.04.049

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, *344*, 109–124. https://doi.org/10.1016/j.heares.2016.11.004

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2018). The association between hearing impairment and neural envelope encoding at different ages. *Neurobiology of Aging*, *74*, 202–212. https://doi.org/10.1016/J.NEUROBIOLAGING.2018.10.008

Goossens, T., Vercammen, C., Wouters, J., & Wieringen, A. Van. (2016). Aging Affects Neural Synchronization to Speech-Related Acoustic Modulations. *Frontiers in Aging Neuroscience*, *8*(June), 1–16. https://doi.org/10.3389/fnagi.2016.00133

Gordon-Salant, S. (2006). Speech Perception and Auditory Temporal Processing Performance by Older Listeners : Implications for Real-World Communication. *Semin Hear*, (27), 264–268. https://doi.org/10.1055/s-2006-954852.

Gosselin, P. A., & Gagné, J.-P. (2011). Older Adults Expend More Listening Effort Than Young Adults Recognizing Speech in Noise. *Journal of Speech, Language, and Hearing Research*, *54*, 944–958. https://doi.org/10.1044/1092-4388(2010/10-0069)

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(November), 887–892.

Grose, J. H., & Mamo, S. K. (2010). Processing of Temporal Fine Structure as a Function of Age. *Ear & Hearing*, *31*(6), 755–760.

Hagerman, B. (1984). Clinical measurements of speech reception thresholds in noise. *Scandinavian Audiology*, *13*(1), 57–63.

Hall, J. W. (2006). *New Handbook of Auditory Evoked Potentials*. Boston: Allyn and Bacon.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hartley, D., Rochtchina, E., Newall, P., Golding, M., & Mitchell, P. (2010). Use of Hearing Aids and Assistive Listening Devices in an Older Australian Population. *Journal of the American Academy of Audiology*, *21*(10), 642–653. https://doi.org/10.3766/jaaa.21.10.4

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. *Neuron*, *57*(3), 452–462. https://doi.org/10.1016/j.neuron.2007.12.009

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hearing Loss Association of America. (2018). Hearing loss statistics.

Heffernan, E., Coulson, N. S., Henshaw, H., Barry, J. G., & Ferguson, M. A. (2016). Understanding the psychosocial experiences of adults with mild-moderate hearing loss : An application of Leventhal's self-regulatory model. *International Journal of Audiology*, *2027*(March), 1–10. https://doi.org/10.3109/14992027.2015.1117663

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, *117*(2), 842–849. https://doi.org/10.1121/1.1836832

Henry, K. S., Kale, S., & Heinz, M. G. (2014). Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Frontiers in Systems Neuroscience*, *8*(February), 1–10. https://doi.org/10.3389/fnsys.2014.00020

Henry, M. J., Herrmann, B., Kunke, D., & Obleser, J. (2017). Aging affects the balance of neural entrainment and top-down neural modulation in the listening brain. *Nature Communications*, *8*(May). https://doi.org/10.1038/ncomms15801

Henshaw, H., Mccormack, A., & Ferguson, M. A. (2015). Intrinsic and extrinsic motivation is associated with computer-based auditory training uptake, engagement, and adherence for people with hearing loss. *Frontiers in Psychology*, *6*(August), 1–13. https://doi.org/10.3389/fpsyg.2015.01067

Hogan, C. A., & Turner, C. W. (1998). High-frequency audibility: benefits for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 104*(1), 432–441. https://doi.org/10.1121/1.423247

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening , parafoveal vision , and visual masking: A survey and appraisal. *The Behavioral and Brain Sciences*, *9,* 1–66.

Holmes, E., Folkeard, P., Johnsrude, I. S., & Scollie, S. (2018). Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *International Journal of Audiology*, *57*(7), 483–492. https://doi.org/10.1080/14992027.2018.1432901

Holmes, E., Kitterick, P. T., & Summerfield, A. Q. (2017). Peripheral hearing loss reduces the ability of children to direct selective attention during multi-talker listening. *Hearing Research*, *350*, 160–172. https://doi.org/10.1016/j.heares.2017.05.005

Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E., & Johnsrude, I. S. (2018). Attentional Modulation of Envelope-Following Responses at Lower (93–109 Hz) but Not Higher (217–233 Hz) Modulation Rates. *JARO - Journal of the Association for Research in Otolaryngology*, *19*(1), 83–97. https://doi.org/10.1007/s10162-017-0641-9

Holube, I., Haeder, K., Imbery, C., & Weber, R. (2016). Subjective Listening Effort and Electrodermal Activity in Listening Situations with Reverberation and Noise. *Trends in Hearing*, *20*, 1–15. https://doi.org/10.1177/2331216516667734

Hopkins, K., & Moore, B. C. J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *The Journal of the Acoustical Society of America*, *122*(1055–1068). https://doi.org/10.1121/1.2749457

Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America*, *123*(2), 1140–1153. https://doi.org/10.1121/1.2824018

Horton, C., Srinivasan, R., & D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a "cocktail party." *Journal of Neural Engineering*, *11*(4). https://doi.org/10.1088/1741-2560/11/4/046015

Howard, M. F., & Poeppel, D. (2010). Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *Journal of Neurophysiology*, *104*, 2500–2511. https://doi.org/10.1152/jn.00251.2010

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kale, S., & Heinz, M. G. (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *JARO - Journal of the Association for Research in*

*Otolaryngology*, *11*(4), 657–673. https://doi.org/10.1007/s10162-010-0223-6

Kale, S., & Heinz, M. G. (2012). Temporal modulation transfer functions measured from auditory-nerve responses following sensorineural hearing loss. *Hearing Research*, *286*(1–2), 64–75. https://doi.org/10.1016/j.heares.2012.02.004

Kaplan, H., & Pickett, J. M. (1982). Differences in speech discrimination in the elderly as a function of type of competing noise: Speech-babble or cafeteria. *Audiology*, *21*, 325–333.

Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*. https://doi.org/10.1098/rstb.2016.0101

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a " Cocktail Party ". *The Journal of Neuroscience*, *30*(2), 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Keurs, M. ter, Festen, J. M., & Plomp, R. (1989). Effects of spectral smearing on speech perception. *The Journal of the Acoustical Society of America*, *86*(S80). https://doi.org/10.1121/1.2027669

Ki, J. J., Kelly, S. P., & Parra, L. C. (2016). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *Journal of Neuroscience*, *36*(10), 3092–3101. https://doi.org/10.1523/JNEUROSCI.2942-15.2016

Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity , uncertainty , and masking in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *111*, 1367–1376. https://doi.org/10.1121/1.1448342

Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., … Petkov, C. I. (2017). *Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex*. *PLoS Biology* (Vol. 15). https://doi.org/10.1371/journal.pbio.2000219

Kochkin, S. (2000). MarkeTrak V : " Why my hearing aids are in the drawer ": The consumers ' perspective. *The Hearing Journal*, *53*(2), 34–41.

Kong, Y. Y., Mullangi, A., & Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, *316*, 73–81. https://doi.org/10.1016/j.heares.2014.07.009

Kortlang, S., Chen, Z., Gerkmann, T., Kollmeier, B., Ewert, S. D., Kortlang, S., … Ewert, S. D. (2018). Evaluation of combined dynamic compression and single channel noise reduction for hearing aid applications. *International Journal of Audiology*, *57*(0), 43–54. https://doi.org/10.1080/14992027.2017.1300695

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*(18), 2867-2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing Aspects of Auditory Handicap by Means of Pupil Dilatation. *International Journal of Audiology*, *36*(3), 155–164. https://doi.org/10.3109/00206099709071969

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503–512. https://doi.org/10.1080/14992020600754583

Krueger, M., Schulte, M., Brand, T., & Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *The Journal of the Acoustical Society of America*, *141*(6), 4680–4693. https://doi.org/10.1121/1.4986938

Kuipers, J. R., & Thierry, G. (2011). N400 Amplitude Reduction Correlates with an Increase in Pupil Size. *Frontiers in Human Neuroscience*, *5*(June), 1–5. https://doi.org/10.3389/fnhum.2011.00061

Kukla, A. (1972). Foundations of an attributional theory of performance. *Psychological Review*, *79*(6), 454–470. https://doi.org/10.1037/h0033494

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.Thirty

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lacher-Fougère, S., & Demany, L. (2005). Consequences of cochlear damage for the detection of interaural phase differences. *The Journal of the Acoustical Society of America*, *118*, 2519–2526. https://doi.org/10.1121/1.2032747

Lankinen, K., Saari, J., Hari, R., & Koskinen, M. (2014). Intersubject consistency of cortical MEG signals during movie viewing. *NeuroImage*, *92*, 217–224. https://doi.org/10.1016/j.neuroimage.2014.02.004

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant. *Journal of the Acoustical Society of America*, *119*, 2445–2454. https://doi.org/10.1121/1.2180210

Lehmann, A., & Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: Relative contributions of frequency and spatial cues. *PLoS ONE*, *9*(1), 1–10. https://doi.org/10.1371/journal.pone.0085442

Loizou, P. P. (2006). Speech Processing in Vocoder-Centric Cochlear Implants. *Adv Otorhinolaryngol.*, *64*, 109–143.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*(April), 1–14. https://doi.org/10.3389/fnhum.2014.00213

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, *103*(49), 18866–18869. https://doi.org/10.1073/pnas.0607364103

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Maddox, R. K., & Lee, A. K. C. (2018). Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. *ENeuro*, *5*(February), 1–13. https://doi.org/10.1523/ENEURO.0441-17.2018

Madsen, S. M. K., Harte, J. M., Elberling, C., & Dau, T. (2018). Accuracy of averaged auditory brainstem response amplitude and latency estimates and latency estimates. *International Journal of Audiology*, *57*, 345–353. https://doi.org/10.1080/14992027.2017.1381770

Magnusson, L., Claesson, A., Persson, M., & Tengstrand, T. (2013). Speech recognition in noise using bilateral open- fit hearing aids: The limited benefit of directional microphones and noise reduction. *International Journal of Audiology*, *52*(1), 29–36. https://doi.org/10.3109/14992027.2012.707335

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech Evoked Potentials: From the Laboratory to the Clinic. *Ear & Hearing*, *29*, 285–313.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95–112. https://doi.org/10.1016/j.jecp.2017.04.006

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology*, *53*, 433–445. https://doi.org/10.3109/14992027.2014.890296

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, *102*(6), 3329. https://doi.org/10.1152/jn.91128.2008.

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG. *Trends in Hearing*, *21*, 1–13. https://doi.org/10.1177/2331216517706396

Millman, R. E., Mattys, S. L., Gouws, A. D., & Prendergast, G. (2017). Magnified Neural Envelope Coding Predicts Deficits in Speech Perception in Noise. *The Journal of Neuroscience*, *37*(32), 7727–7736. https://doi.org/10.1523/JNEUROSCI.2722-16.2017

Millman, R. E., Prendergast, G., Hymers, M., & Green, G. G. R. (2013). Representations of the temporal envelope of sounds in human auditory cortex: Can the results from invasive intracortical "depth" electrode recordings be replicated using non-invasive MEG "virtual electrodes"? *NeuroImage*, *64*(1), 185–196. https://doi.org/10.1016/j.neuroimage.2012.09.017

Moon, I. J., & Hong, S. H. (2014). What is temporal fine structure and why is it important ? *Korean Journal of Audiology*, *18*(1), 1–7.

Moore, B. C. J. (1995). *Perceptual consequences of cochlear damage*. (O. U. Press, Ed.). Oxford.

Moore, B. C. J. (2008). The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People. *Journal of the Association for Research in Otolaryngology*, *406*, 399–406. https://doi.org/10.1007/s10162-008-0143-x

Moore, B. C. J., & Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *The Journal of the Acoustical Society of America*, *94*, 2050–2062. https://doi.org/10.1121/1.407478

Moore, B. C. J., Glasberg, B. R., & Vickers, D. A. (1999). Further evaluation of a model of loudness perception applied to cochlear hearing loss. *The Journal of the Acoustical Society of America*, *106*, 898–907. https://doi.org/10.1121/1.427105

Moore, B. C. J., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *The Journal of the Acoustical Society of America*, *105*(1), 400–411. https://doi.org/10.1121/1.424571

Moore, B. C. J., Wojtczak, M., & Vickers, D. A. (1996). Effect of loudness recruitment on the perception of amplitude modulation. *The Journal of the Acoustical Society of America*, *100*, 481–489. https://doi.org/10.1121/1.415861

Moore, T. M., Key, A. P., Thelen, A., & Hornsby, B. W. Y. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia*, *106*(May), 371–382. https://doi.org/10.1016/j.neuropsychologia.2017.10.025

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Mulert, C., Seifert, C., Leicht, G., Kirsch, V., Ertl, M., Karch, S., … Jäger, L. (2008). Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *NeuroImage*, *42*(1), 158–168. https://doi.org/10.1016/j.neuroimage.2008.04.236

Musacchia, G., Ortiz-Mantilla, S., Roesler, C. P., Rajendran, S., Morgan-Byrne, J., & Benasich, A. A. (2018). Effects of noise and age on the infant brainstem response to speech. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2018.08.005

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing. *International Journal of Audiology*, *48*(10), 684–691. https://doi.org/10.1080/14992020902962421

Nilsson, M., Gelnett, D., Sullivan, J., Soli, S. D., & Goldberg, R. L. (1992). Norms for the hearing in noise test: The influence of spatial separation, hearing loss, and English language experience on speech reception thresholds. *The Journal of the Acoustical Society of America*, *92*(4), 2385–2385. https://doi.org/10.1121/1.404787

Nuttall, H. E., Heinrich, A., Moore, D., & De Boer, J. (2013). Do cochlear mechanisms explain the noise-disruption of the auditory brainstem response to speech? *Proceedings of Meetings on Acoustics*, *19*. https://doi.org/10.1121/1.4792844

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., … Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Oberem, J., Koch, I., & Fels, J. (2017). Intentional switching in auditory selective attention: Exploring age-related effects in a spatial setup requiring speech perception. *Acta Psychologica*, *177*, 36–43.

https://doi.org/10.1016/j.actpsy.2017.04.008

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., … Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing*, *38*(3), 267–281. https://doi.org/10.1097/AUD.0000000000000396

Olguin, A., Bekinschtein, T. A., & Bozic, M. (2018). Neural Encoding of Attended Continuous Speech under Different Types of Interference. *Journal of Cognitive Neuroscience, 30*(11), 1606–1619. https://doi.org/10.1162/jocn

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. In Neuper & Klimesch (Ed.), *Progress in Brain Research*.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip : Open Source Software for Advanced Analysis of MEG , EEG , and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2011/156869

Oxenham, A. J. (2018). How we hear : The perception and neural coding of sound. *Annual Review of Psychology*, *69*, 27–50.

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia*, *49*(12), 3338–3345. https://doi.org/10.1016/j.neuropsychologia.2011.08.007

Parbery-Clark, Alexandra, Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear & Hearing*, *30*(6), 653–661.

Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, *78*(3), 321–338. https://doi.org/10.1007/s00426-013-0534-4

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*(SEP), 1–17. https://doi.org/10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*, 1378–1387. https://doi.org/10.1093/cercor/bhs118

Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cerebral Cortex*, *20*(4), 773–782. https://doi.org/10.1093/cercor/bhp142

Peissig, J., & Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, *101*(3), 1660–1670. https://doi.org/10.1121/1.418150

Peng, S. L., Chen, X., Li, Y., Rodrigue, K. M., Park, D. C., & Lu, H. (2018). Age-related changes in cerebrovascular reactivity and their relationship to cognition: A four-year longitudinal study. *NeuroImage*, *174*(March), 257–262. https://doi.org/10.1016/j.neuroimage.2018.03.033

Petersen, E. B., Wöstmann, M., Obleser, J., & Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology*, *117*(1), 18–27. https://doi.org/10.1152/jn.00527.2016

Petroni, A., Cohen, S. S., Ai, L., Langer, N., Henin, S., Vanderwal, T., … Parra, L. C. (2018). The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. *Eneuro*, *5*(February), ENEURO.0244-17.2017. https://doi.org/10.1523/ENEURO.0244-17.2017

Pichora-Fuller, M. K. (2016). How social psychological factors may modulate auditory and cognitive functioning during listening. *Ear & Hearing*, *37*, 92–100.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., … Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and*

*Hearing*, *37*(Supplement 1), 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Picou, E., Aspell, E., & Ricketts, T. (2014). Potential Benefits and Limitations of Three Types of Directional Processing in Hearing Aids. *Ear Hear*, *35*(3), 339–352. https://doi.org/10.1097/AUD.0000000000000004

Picou, E. M., Moore, T. M., & Ricketts, T. A. (2017). The effects of directional processing on objective and subjective listening effort. *Journal of Speech, Language, and Hearing Research*, *60*, 199–211. https://doi.org/10.1044/2016

Picou, E. M., & Ricketts, T. A. (2014). Increasing motivation changes subjective reports of listening effort and choice of coping strategy. *International Journal of Audiology*, *53*(6), 418–426. https://doi.org/10.3109/14992027.2014.880814

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing-impaired. *Journal of Speech, Language, and Hearing Research*, *29*(June), 146–154.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*, 1–9. https://doi.org/10.1038/srep43916

Presacco, A., Simon, J. Z., & Anderson, S. (2016a). Effect of informational content of noise on speech representation in the aging midbrain and cortex. *Journal of Neurophysiology*, *116*(5), 2356–2367. https://doi.org/10.1152/jn.00373.2016

Presacco, A., Simon, J. Z., & Anderson, S. (2016b). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *Journal of Neurophysiology*, *116*(5), 2346–2355. https://doi.org/10.1152/jn.00372.2016

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual Organization of Sound Begins in the Auditory Periphery. *Current Biology, 18*(15), 1124–1128. https://doi.org/10.1016/j.cub.2008.06.053

Prosser, S., Turrini, M., & Arslan, E. (1991). Effects of different noises on speech discrimination by the elderly. *Acta Oto-Laryngol. Suppl., 476,* 136–142.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114,* 446–454. https://doi.org/10.1121/1.1579009

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope vocoder processing on F0 discrimination and concurrent vowel identification. *Ear and Hearing*, *26*, 451–460.

R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna.

Reichenbach, C. S., Braiman, C., Schiff, N. D., Hudspeth, A. J., & Reichenbach, T. (2016). The Auditory-Brainstem Response to Continuous, Non-repetitive Speech Is Modulated by the Speech Envelope and Reflects Speech Processing. *Frontiers in Computational Neuroscience*, *10*(May), 47. https://doi.org/10.3389/fncom.2016.00047

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, *118*, 1274–1277. https://doi.org/10.1121/1.2000751

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2008). Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing. *Ear and Hearing*, *29*(2), 169–175. https://doi.org/10.1097/AUD.0b013e31816476d4

Richter, M. (2013). A Closer Look Into the Multi-Layer Structure of Motivational Intensity Theory. *Social and Personality Psychology Compass*, *1*(7), 1–12.

Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, *37*, 111S-117S. https://doi.org/10.1097/AUD.0000000000000295

Ricketts, T. A., & Hornsby, B. W. Y. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing “digital noise reduction.”

*Journal of the American Academy of Audiology*, *16*(5), 270–277. https://doi.org/10.3766/jaaa.16.5.2

Ricketts, T. A., & Hornsby, B. W. Y. (2006). Directional hearing aid benefit in listeners with severe hearing loss. *International Journal of Audiology*, *45*(3), 190–197. https://doi.org/10.1080/14992020500258602

Ricketts, T. A., Picou, E. M., & Galster, J. (2017). Directional microphone hearing aids in school environments: Working toward optimization. *Journal of Speech, Language, and Hearing Research*, *60*, 263–275. https://doi.org/10.1044/2016

Robles, L., & Ruggero, M. A. (2001). Mechanics of the Mammalian Cochlea. *Physiological Reviews*, *81*(3), 1305–1352.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences*, *336*(1278), 367–373. Retrieved from http://www.jstor.org/stable/55906

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443. https://doi.org/10.1121/1.4794379

Russell, E. L. (2015). *Auditory and Visual Sustained Attention on Tasks with Varied Motivation and Cognitive Loads in Children With and Without ADHD*.

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, *156*(1), 95–103. https://doi.org/10.1016/j.bbr.2004.05.012

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*(9), 2021–2030. https://doi.org/10.1016/j.clinph.2004.04.003

Santurette, S., & Dau, T. (2007). Binaural pitch perception in normal-hearing and hearing-impaired listeners. *Hearing Research*, *223*, 29–47. https://doi.org/10.1016/j.heares.2006.09.013

Sek, A., Baer, T., Crinnion, W., Springgay, A., Moore, B. C. J., Baer, T., … Springgay, A. (2015). Modulation masking within and across carriers for subjects with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *138*, 1143–1153. https://doi.org/10.1121/1.4928135

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., & Best, V. (2015). Selective Attention in Normal and Impaired Hearing. *Cortex*, *68*, 144–154. https://doi.org/10.1177/1084713808325306

Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *Journal of the American Academy of Audiology*, *26*(2), 183–196. https://doi.org/10.3766/jaaa.26.2.7

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Letters to Nature, 416*, 87–90.

Solheim, J., Kværner, K. J., & Falkenberg, E. (2011). Daily life consequences of hearing loss in the elderly. *Disability and Rehabilitation*, *33*(23–24), 2179–2185. https://doi.org/10.3109/09638288.2011.563815

Song, J. H., Nicol, T., & Kraus, N. (2011). Test-retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, *122*(2), 346–355. https://doi.org/10.1016/j.clinph.2010.07.009

Song, J., & Iverson, P. (2018a). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*(June), 163–170. https://doi.org/10.1016/j.cognition.2018.06.001

Song, J., & Iverson, P. (2018b). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*(November 2016), 163–170. https://doi.org/10.1016/j.cognition.2018.06.001

Sörös, P., Teismann, I. K., Manemann, E., & Lütkenhöner, B. (2009). Auditory temporal processing in healthy aging: A magnetoencephalographic study. *BMC Neuroscience, 10*, 1–9. https://doi.org/10.1186/1471-2202-10-34

Southall, K., Gagné, J., & Jennings, M. B. (2010). Stigma: A negative and a positive influence on help-seeking for adults with acquired hearing loss. *International Journal of Audiology*, *49*(11), 804–814. https://doi.org/10.3109/14992027.2010.498447

Steinschneider, M., & Dunn, M. (2002). Electrophysiology in developmental neuropsychology. In S. Segalowitz & I. Rapin (Eds.), *Handbook of neuropsychology* (2nd ed., pp. 91–146). Amsterdam: Elsevier.

Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special? *Hearing Research*, *305*(1), 57–73. https://doi.org/10.1016/j.heares.2013.05.013

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, *116*(2), 1081–1091. https://doi.org/10.1121/1.1772399

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1–2), 22–29. https://doi.org/10.1016/j.heares.2009.12.021

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3), 191–201. https://doi.org/10.1016/j.bandl.2012.09.001

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed Expectancies under Degraded Speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. https://doi.org/10.1162/jocn

Stringer, L. (2015). *Accent intelligibility across native and non-native accent pairings: Investigating links with electrophysiological measures of word recognition*. University College London.

Summers, V., & Leek, M. R. (1994). The internal representation of spectral contrast in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *95*(6), 3518–3528. https://doi.org/10.1121/1.409969

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality (Sussman, Winkler et Schröger, 2003), *10*(3), 630–637.

Tabachnick, A., & Toscano, J. (2018). Perceptual Encoding in Auditory Brainstem Responses: Effects of Stimulus Frequency. *Journal of Speech, Language, and Hearing Research*, 1–12.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (Eds.), *Speech Coding and Synthesis*. New York: Elsevier.

Taylor, W. L. (1953). “Cloze Procedure”: A New Tool for Measuring Readability. *Journalism Bulletin*, *30*(4), 415–433. https://doi.org/10.1177/107769905303000401

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*(December), 1–7. https://doi.org/10.3389/fpsyg.2013.00855

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *PNAS*, *112*(32), 10062–10067. https://doi.org/10.1073/pnas.1505114112

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual Lexical Access During L1 Sentence Reading: The Effects of L2 Knowledge , Semantic Constraint , and L1 – L2 Intermixing. *Journal of Experimental Psychology*, *37*(6), 1412–1431. https://doi.org/10.1037/a0024492

Tremblay, K., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical*

*Neurophysiology*, *114*, 1332–1343. https://doi.org/10.1016/j.jcomdis.2007.03.008

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, Hearing Acuity, and the Attentional Costs of Effortful Listening. *Psychology and Aging*, *24*(3), 761–766. https://doi.org/10.1037/a0014802

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453–467. https://doi.org/10.1037/0882-7974.17.3.453

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *JARO - Journal of the Association for Research in Otolaryngology*, *19*(2), 181–191. https://doi.org/10.1007/s10162-018-0654-z

Varghese, L., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, *1626*, 146–164. https://doi.org/10.1016/j.brainres.2015.06.038

Varnet, L., Ortiz-Barajas, M. C., Guevara Erra, R., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, *142*, 1976–1989. https://doi.org/10.1121/1.5006179

Verleger, R. (1988). Event-related potentials and cognition: a critique of the context updating hypothesis and an alternative explanation of P3. *Behav. Brain Sci.*, *11*, 343–356.

Vermeire, E., Hearnshaw, H., Van Royen, P., & Denekens, J. (2001). Patient adherence to treatment : three decades of research . A comprehensive review. *Journal of Clinical Pharmacy and Therapeutics*, *26*, 331–342.

Walden, B. E., Surr, R. K., Cord, M. T., Edwards, B., & Olson, L. (2000). Comparison Of Benefits Provided By Different Hearing Aid Technologies. *Journal of the American Academy of Audiology*, *11*(10), 540–560. https://doi.org/10.1007/s12348-011-0058-2

Wallaert, N., Moore, B. C. J., Ewert, S. D., & Lorenzi, C. (2017). Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation. *The Journal of the Acoustical Society of America*, *141*, 971–980. https://doi.org/10.1121/1.4976080

Wallhagen, M. I. (2009). The Stigma of Hearing Loss. *The Gerontologist*, *50*(1), 66–75. https://doi.org/10.1093/geront/gnp107

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., … Gershon, R. C. (2013). Vision assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S37–S40. https://doi.org/10.1212/WNL.0b013e3182876e0a

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of Noise and Noise Reduction on Processing Effort. *Ear and Hearing*, *38*(6), 690–700. https://doi.org/10.1097/AUD.0000000000000454

WHO. (2018). Deafness and hearing loss. Retrieved from http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230–242. https://doi.org/10.1093/cercor/bhm049

Winn, M., Edwards, J. R., & Litovsky, R. Y. (2013). The impact of spectral resolution on listening effort revealed by pupil dilation. *The Journal of the Acoustical Society of America*, *134*(5), 4233–4233. https://doi.org/10.1121/1.4831554

Woestmann, M., Lim, S.-J., & Obleser, J. (2017). The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. *Cerebral Cortex*, *27*(June), 3307–3317. https://doi.org/10.1093/cercor/bhx074

Wong, L. L. N., Ng, E. H. N., & Soli, S. D. (2012). Characterization of speech understanding in various types of noise. *The Journal of the Acoustical Society of America*, *132*(4), 2642–2651. https://doi.org/10.1121/1.4751538

Wood, N., & Cowan, N. (1995). The Cocktail Party Phenomenon Revisited: How Frequent Are Attention Shifts to One's Name in an Irrelevant Auditory

Channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260. https://doi.org/10.3892/mmr.2015.4344

Wright, R. A. (1996). Brehm's theory of motivation as a model of effort and cardiovascular response. In *Brehm's theory of motivation as a model of effort and cardiovascular response* (pp. 424–453).

Wright, R. A. (2008). Refining the Prediction of Effort : Brehm ' s Distinction between Potential Motivation and Motivation Intensity. *Social and Personality Psychology Compass*, *2*(2), 682–701.

Wright, R. A., & Franklin, J. (2004). Ability perception determinants of effort-related cardiovascular response: Mood, optimism, and performance resources. In R. A. Wright, J. Greenberg, & S. S. Brehm (Eds.), *Motivational Analyses of Social Behavior: Building on Jack Brehm's Contributions to Psychology*. Hillsdale, NJ: Erlbaum.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology,* 459–482. https://doi.org/10.1037/h0073415

Zhang, C., Arnott, S. R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X., … Schneider, B. A. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hearing Research, 331*, 119–130. https://doi.org/10.1016/j.heares.2015.11.002

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron*, *77*, 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Zoefel, B., & VanRullen, R. (2015a). Selective Perceptual Phase Entrainment to Speech Rhythm in the Absence of Spectral Energy Fluctuations. *Journal of Neuroscience*, *35*(5), 1954–1964. https://doi.org/10.1523/JNEUROSCI.3484-14.2015

Zoefel, B., & VanRullen, R. (2015b). The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Frontiers in Human Neuroscience*, *9*, 1–12. https://doi.org/10.3389/fnhum.2015.00651