UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Psychologie I
der Universität zu Lübeck

Direktorium: Prof. Dr. Nico Bunzeck, Prof. Dr. Jonas Obleser, Prof. Dr. Corinna Peifer

Selective attention in multi-talker situations: neural and behavioral mechanisms

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

vorgelegt von
Martin Orf
aus Bad Hersfeld

Lübeck, 2023

1. Berichterstatter/Berichterstatterin:

Prof. Dr. Jonas Obleser

2. Berichterstatter/Berichterstatterin:

Prof. Dr. Jonathan Simon

Tag der mündlichen Prüfung:

02. August 2023

Zum Druck genehmigt. Lübeck, den

24. Oktober 2023

Selective attention in multi-talker situations: neural and behavioral mechanisms

Contents

1 General introduction 4
1.1 Auditory object formation . . . 5
1.2 Auditory selective attention . . . 6
1.2.1 Early versus late selection . . . 6
1.2.2 Bottom-up versus top-down attention . . . 7
1.2.3 Exploring attentional background through the principles of negative priming in a continuous speech paradigm . . . 8
1.2.4 Neutral baseline for distractor suppression investigation . . . 9
1.2.5 Neural signatures of selective attention and neural speech tracking . . . 9
1.3 Implications of presbycusis on the peripheral hearing and selective attention . . . 12
1.3.1 Object formation and selective attention in hearing impaired listeners . . . 12
1.3.2 Changes along the auditory pathway due to presbycusis . . . 13
1.3.3 Neural speech tracking in hearing impaired listeners . . . 15
1.3.4 Hearing aid algorithm . . . 15
1.4 Research questions . . . 17

2 General methods 18
2.1 Psychophysically augmented continuous speech stimuli . . . 18
2.2 Electroencephalograpy . . . 20
2.2.1 EEG data preprocessing . . . 20
2.3 Feature extraction from continuous speech . . . 21
2.4 Neural response to continuous stimuli . . . 22
2.4.1 Encoding versus decoding models . . . 23
2.4.2 Temporal response function . . . 25
2.4.3 Regularization . . . 25
2.4.4 Training and testing . . . 26
2.4.5 Neural tracking . . . 27
2.5 Dynamic range compression . . . 27
2.6 Peripheral auditory modelling . . . 30
2.7 Statistical analysis . . . 32

3 Study 1: Auditory neural tracking reflects target enhancement but not distractor suppression in a psychophysically augmented continuous-speech paradigm 36
3.1 Abstract . . . 36
3.2 Introduction . . . 36
3.3 Methods . . . 39
3.3.1 Participants . . . 39
3.3.2 Stimulus materials and spatial cue . . . 39
3.3.3 Experimental Setup . . . 40
3.3.4 Behavioural data analysis . . . 41
3.3.5 Data acquisition and pre-processing . . . 42
3.3.6 Extraction of the speech envelope . . . 43
3.3.7 Temporal response functions (TRFs) . . . 43
3.3.8 Neural tracking . . . 44
3.3.9 Statistical analysis . . . 44
3.3.10 Statistical analysis on time series . . . 45
3.4 Results . . . 45
3.4.1 Larger repeat evoked responses in the target stream . . . 46
3.4.2 Larger interference by distracting versus neutral speech . . . 47
3.4.3 Morphology of neural responses to target, neutral, and distractor speech . . . 48
3.4.4 Neural tracking reflects target enhancement, not distractor suppression . . . 50
3.4.5 Neural tracking of the target stream is associated with perceptual performance . . . 50
3.4.6 Control Analysis I: Listeners process the content of competing speech streams . . . 52
3.4.7 Control Analysis II: Condition-to-location assignment does not confound interference by distracting speech and sub-processes of attention . . . 52
3.4.8 Control Analysis III: Unfolding of neural filters (TRFs) across trial duration . . . 53

3.5 Discussion . . . 53
3.5.1 Neural tracking of speech implements enhancement, not suppression . . . 54
3.5.2 Auditory attention exploits statistical regularities to separate distracting versus neutral speech . . . 56
3.5.3 Neural tracking of target but not distractor explains performance . . . 58
3.6 Limitations . . . 58
3.7 Conclusion . . . 59

4 Exploring the interplay between dynamic range compression and selective attention in competing-talker environments 60
4.1 Introduction . . . 60
4.2 Methods . . . 63
4.2.1 Normal hearing participants . . . 63
4.2.2 Hearing impaired participants . . . 63
4.2.3 Stimulus materials . . . 64
4.2.4 Experimental setup . . . 65
4.2.5 Experimental procedure . . . 66
4.2.6 Sound pressure level adjustment for hearing impaired participants . . . 67
4.2.7 Data acquisition and pre-processing . . . 68
4.2.8 Extracting the speech envelope . . . 68
4.2.9 Temporal response function and neural tracking estimation . . . 68
4.2.10 Statistical analysis . . . 69
4.2.11 Statistical analysis on time series . . . 70
4.3 Study 2: The effects of compression ratio on neural tracking of speech: A pilot study . . . 70
4.4 Study 3: The interplay between dynamic range compression and selective attention in normal hearing listeners . . . 73
4.4.1 Compression on both streams impairs performance . . . 73
4.4.2 Compression decreases neural speech tracking . . . 76
4.4.3 Increased neural separation due to decreased neural tracking of ignored stream . . . 78
4.5 Study 4: Auditory peripheral modeling of compressed and uncompressed speech . . . 80
4.6 Study 5: The interplay between dynamic range compression and selective attention in hearing impaired listeners . . . 82
4.6.1 Slower responses to attended compressed speech . . . 83
4.6.2 Hearing impaired show stronger neural tracking than normal hearing . . . 84
4.6.3 Compression decreases neural tracking in hearing impaired . . . 85
4.6.4 Largest neural separation in the attend uncompressed - ignore compressed attentional pair . . . 86
4.6.5 Hearing impaired show larger differences between attended and ignored streams compared to normal hearing . . . 89
4.6.6 Control analysis: Front back location assignment does not confound neural and behavioural results . . . 89
4.7 Study 6: perceived loudness of compressed and uncompressed speech stimuli . . . 90
4.7.1 Participants . . . 91
4.7.2 Methods . . . 91
4.7.3 No significant differences in perceived loudness between uncompressed and Zwicker loudness matched compressed speech . . . 92
4.8 Discussion . . . 94
4.8.1 Amplitude compression on both attend and ignore speech impairs performance . . . 94
4.8.2 Amplitude compression on ignore speech may facilitate inhibition . . . 95
4.8.3 Unexpected lack of effects of amplitude compression applied to single speech streams 96
4.8.4 Effects of amplitude compression on performance for hearing impaired participants 97
4.8.5 Amplitude compression decreases neural tracking of speech . . . 97
4.8.6 Neural tracking and selective attention: Effects of compression and attentional pairs 98
4.8.7 Hearing impaired participant showed larger neural tracking responses compared to normal hearing participants . . . 99
4.9 Limitations . . . 101
4.10 Conclusion . . . 102

5 General discussion **103**
5.1 Summary of experimental results 103
5.2 Terminological considerations 104
5.3 The role of short repeats as a behavioural measure for assessing selective attention in continuous speech 105
5.4 Implication of the psychophysically augmented continuous speech paradigm on stream formation 107
5.5 Investigating the sub-mechanisms of selective attention: Intersection of findings from two parts of the thesis 109
5.6 Implications on hearing aids and future research 110
5.7 Limitations 112

6 Conclusions **114**

References **115**

Summary **130**

Zusammenfassung **132**

Curriculum vitae **135**

1 General introduction

In everyday life, people are often confronted with complex auditory environments in which multiple sound sources compete for their attention. The ability to selectively attend to one particular sound source over others is essential for effective communication, and this is commonly referred to as the ”cocktail party problem” (Cherry, 1953). While the healthy auditory system is remarkably adept at solving such complex auditory scenes, even mild to moderate hearing loss can impair the processing of speech in complex environments. To better understand the neural mechanisms of selective attention, my first study investigated these mechanisms in young, healthy participants using a psychophysically augmented continuous-speech paradigm. In the second part of the thesis, I investigate the relationship between attention and amplitude compression in young and older participants with mild to moderate sensorineural hearing loss. The goal of this research is twofold: to enhance our understanding of neural mechanisms underlying selective attention, and to lay a foundation to support hearing-impaired listeners in improving their ability to handle complex auditory environments.

1.1 Auditory object formation

To analyse the auditory scene, a listener must separate one or multiple sound sources by creating an internal representation of them, referred to as ”auditory objects” (Bregman, 1994). Although auditory objects play an important role in selective auditory attention in the auditory scene, the precise definition of an object is not always straightforward in audition, as the listener’s expectations and state also determine the constitution of an auditory object. For instance, listening to the entire symphony orchestra versus listening to a single instrument. Although the definition of an auditory object is not quite simple, sound emitted from a single physical sound source usually refers to the perceptual entity of an auditory object (B. G. Shinn-Cunningham, 2008). According to a more formal definition, auditory objects are the computational output of the auditory system’s ability to locate, isolate, classify, and organise the spectrotemporal regularities present in the acoustic environment into stable perceptual units (Bizley & Cohen, 2013).

In the cocktail-party scenario, continuous speech is usually the physical sound source of interest. But what is needed to prioritise relevant speech over distracting speech and thus solve the cocktail party problem? The listener’s ability to solve the cocktail party problem relies on the spectrotemporal structure of speech, which is important on different time scales. Auditory object formation and auditory selective attention are two processes associated with solving the cocktail party problem.

The structure of speech has different levels of importance depending on the time scale. Object formation operates at two time scales: a local time scale that binds sound energy spectrotemporally linked to short-term objects (integration of simultaneous components), and a longer scale that connects these short-term objects into auditory objects that evolve over time (integration of sequential components). Bregman (1978) referred to an auditory object that extends in time, such as syllables linked together over time, as a ”stream.” In the following, I will use the term ”stream” for an auditory object composed of multiple syllables.

Computing relevant feature values in each syllable is key to successful streaming. For instance, frequency, pitch, timbre, amplitude modulation rate, and spatial location of sequentially presented syllables all contribute to the perception of the stream as one single continuous source (B. G. Shinn-Cunningham, 2008; Shamma, Elhilali, & Micheyl, 2011; Griffiths & Warren, 2004). It has been suggested that the neural representations of auditory objects are formed in the auditory cortex by combining all the features that belong to the same object. These representations are considered to be basic units for higher-level cognitive processes (Snyder, Gregg, Weintraub, & Alain, 2012; Nelken & Bar-Yosef, 2008). Listeners perceive continuous speech as one auditory stream despite the presence of silent gaps where spectrotemporal continuity is absent. Therefore, higher-order perceptual features are also associated with auditory streaming (B. Shinn-Cunningham, Best, & Lee, 2017).

In the present thesis, I primarily focus on the attentional processing of different talkers presented at different spatial positions. Regardless of which cue dominates object formation, I primarily investigate the neural response to the broad-band temporal dynamics (represented as the envelope of the speech signal) of the talkers.

1.2 Auditory selective attention

In the cocktail party scenario, listeners have to select which stream they want to attend since it is impossible to process every talker in detail due to limited processing capacity (Marois & Ivanoff, 2005). In addition, it is usually the listener's goal to communicate with a single talker rather than to analyse the complete auditory scene. Selective attention refers to listeners' cognitive ability to control which information they want to attend to (object selection) in the simultaneous presence of distractors (Desimone, Duncan, et al., 1995). Attention is focused on objects even when observers choose what to pay attention to based on low-level features. When attention is spatially focused, observers become more sensitive to additional features that are specific to the attended object (Duncan, 2006). As a result, object formation and selective attention are closely related, with the object serving as the unit of attention. (B. G. Shinn-Cunningham, 2008). In addition, the temporal coherence theory assumes that attention plays a key role not only in object selection but also in object formation. That is, starting the binding process, altering the neuronal representations of the acoustic features and/or the temporal coherence patterns between these features, and generating the binding signal (Shamma et al., 2011). In challenging listening situations, a failure of object formation could thus be associated with impaired selective attention. However, it was also shown that attention is not always required for the formation of auditory objects (Micheyl et al., 2003). But selective attention can affect how aware we are of an object (B. G. Shinn-Cunningham, 2008). According to the "Load Theory of Attention" proposed by (Lavie, 1995), successful selection in the visual domain depends also on the perceptual demands imposed by the relevant task information. However, in the auditory domain, findings regarding the role of perceptual load on attentional selection have been mixed (Murphy, Spence, & Dalton, 2017). Dichotomous concepts like early versus late and bottom-up versus top-down attention, which are described in the following, are closely related to selective attention.

1.2.1 Early versus late selection

In attentional research, there has been a longstanding debate about whether the cognitive system filters out to-be-ignored stimuli early or late in the processing stages. Early selection theorists, such as Broadbent (1958), propose that the processing of incoming data is limited by computational resources, and that a filter is tuned to reduce the amount of information coming in. Later, Deutsch and Deutsch (1963) proposed that the filter acts only after semantic analysis of all stimuli, thus suggesting late selection.

Early selection implies that irrelevant stimuli are rejected at an early stage based on basic physical characteristics such as location or pitch, rather than their content. However, evidence in favour of late-selection theories indicates that irrelevant stimuli can undergo semantic analysis. In a dichotic listening experiment, Moray (1959) found that participants could still recognise their own names in the unattended stream, suggesting that irrelevant stimuli sometimes undergo semantic analysis. Deutsch and Deutsch (1963) model even presupposes that all input is processed up to the semantic representation and selection functions at this stage.

Based on evidence for both early and late selection, Treisman (1960) proposed the ”attenuation theory” to account for how unattended stimuli could occasionally be processed. This theory advanced Broadbent’s original early selection theory by contending that attenuation occurs on the basis of physical features rather than complete rejection. As long as unattended stimuli still have enough ”strength” after attenuation to pass through a hierarchical analysis process, it would be challenging but not impossible to extract meaningful content from irrelevant inputs.

More recent theories suggest that the extent of attentional selection, whether it occurs early or late, may not be predetermined and may depend on factors such as the task demand and the amount of perceptual load that needs to be processed (Lavie, 1995; Huang-Pollock, Carr, & Nigg, 2002).

1.2.2 Bottom-up versus top-down attention

Bottom-up attention refers to attention that is guided solely by external stimuli that are salient due to their inherent properties relative to the background, while top-down attention refers to attentional guidance based on prior knowledge, deliberate plans, and current goals (e.g., Awh, Belopolsky, & Theeuwes, 2012; Connor, Egeth, & Yantis, 2004; B. G. Shinn-Cunningham & Best, 2008).

A prominent example of bottom-up attention is the sound of one’s own name in ignored speech in the cocktail party. The sound of one’s own name can capture attention even when it emerges in the unattended background (e.g., Conway, Cowan, & Bunting, 2001; Moray, 1959; Holtze, Jaeger, Debener, Adiloğlu, & Mirkovic, 2021). Although attentional capture is stimulus-driven, salience is not pre-programmed in the auditory system but can be developed through learning. In addition, recent behavioural investigations have demonstrated that learning based on prior experiences with distracting information influences the ability to ignore distracting information by extracting statistical regularities from a particular feature over time, such as the position of a distractor (Wang & Theeuwes, 2018b; Daly & Pitt, 2021).

A typical top-down goal is to understand one’s conversation partner in a complex listening situation. When the listener knows that the conversation partner (object) has desired features such as pitch or location, their neural representation becomes more inclined to prioritize those objects over others that don’t have the desired feature (e.g., B. Shinn-Cunningham et al., 2017). The neural signatures of selective attention are described in section 1.3.5.

Mainly in the visual domain, the strong interplay between bottom-up and top-down attention was shown (Egeth & Yantis, 1997). Due to the strong interplay, it is challenging to clearly separate bottom-up and top-down attention using experimental paradigms (e.g., Shuai & Elhilali, 2014). Recent visual studies, however, have shown that top-down inhibitory mechanisms can be used to inhibit bottom-up capture, potentially bridging the gap between bottom-up and top-down theories (Gaspelin, Leonard, & Luck, 2015, 2017). According to the signal suppression hypothesis, salient objects automatically generate a priority signal that capture attention, which is in line with bottom-up theories, but the salient items can be suppressed before, which is in line with top-down theories (Gaspelin & Luck, 2018). Closely related are the two principles of proactive and reactive suppression (van Moorselaar & Slagter, 2020; Geng, 2014). Proactive mechanisms suppress distractions before they occur and reactive mechanisms suppress distraction after the distractor captures attention. Thus, attentional capture can be considered a prerequisite for reactive suppression (Gaspelin & Luck, 2018).

1.2.3 Exploring attentional background through the principles of negative priming in a continuous speech paradigm

Until now, investigating the attentional background in attentional selection based on behavioural measurements has been difficult, because asking listeners about their perceptions of the background would shift the attentional background into the foreground. An influential paradigm in cognitive psychology, negative priming, was first created to address this problem by measuring attentional selective inhibition of distracting information.

Negative priming refers to a subconscious memory and inhibition phenomenon in which a previously presented stimulus leads to slow and error-prone responses to the currently presented stimulus (Tipper, 1985). Classical negative priming designs consist of two main components: the prime (trial N) and the probe (trial N+1). A target is presented together with a distractor in both the prime and the probe. The prime presents a certain stimulus (or stimulus feature) as a distractor, which becomes the target in the probe trial (distractor-to-target repetition).

In the literature, there are primarily two potential underlying processes of negative priming described: inhibition and retrieval mechanisms. The inhibition theory assumes that the distractor is actively suppressed by mechanisms of selective attention throughout the processing of the prime, and that this inhibition lasts until the presentation of the following probe (Houghton & Tipper, 1984; Tipper, 1985). On the other hand, retrieval theories assume that perceiving a target in probe activates memory traces associated with this stimulus in the prime (Neill, Valdes, Terry, & Gorfein, 1992). In other words, the current target still holds information associated with the previously presented distractor, like ”ignore the distractor”. This theory is related to the concept of ”event files,” which postulates that information about the stimulus and response information get incorporated into ”event files” (Hommel, 1998). At present, most researchers agree that both the inhibition mechanism and retrieval processes contribute to negative priming (for review, see Frings, Schneider, & Fox, 2015).

Negative priming evidence was primarily investigated in the visual modality (for review, see E. Fox, 1995). However, negative priming was observed in the auditory modality as well (e.g., Banks, Roberts,

& Ciranni, 1995; Buchner & Mayr, 2004). The general rules of auditory negative priming are similar to those of its visual counterpart (for review, see Mayr & Buchner, 2007). Given the differences between auditory and visual attention, this is by no means a simple matter. For instance, the involvement of peripheral mechanisms such as eye and head movements supports selective attention in the visual domain. In contrast, an equivalent peripheral movement is not present in the auditory domain. Importantly, a more recent study found evidence for negative priming in an auditory selective attention switching task that used speech as stimuli and varied not only the identity of the talker also the location (Eben, Koch, Jolicoeur, & Nolden, 2020).

In this thesis, instead of using a traditional negative priming experiment with a prime and probe structure, I employed the principles of negative priming within a continuous speech paradigm. The details of this approach are explained in Section 3.

1.2.4 Neutral baseline for distractor suppression investigation

The mechanism of how selective attention is implemented at the neural level is still an ongoing debate in attention research. To separate target enhancement and distractor suppression, the two sub-processes of selective attention, it is argued that a pre-defined baseline is needed to test whether the target exceeds the baseline (enhancement) and the distractor falls below the baseline (suppression). Due to the sole separation between the target and distractor, it is insufficient to distinguish between these attentional subprocesses. (Forschack, Gundlach, Hillyard, & Müller, 2022; Wöstmann et al., 2022). The aforementioned studies usually lack such a neutral control condition.

However, in visual attentional research, such a control baseline was successfully implemented. Seidl, Peelen, and Kastner (2012) measured brain activity in response to photographs that contained objects from task-relevant (target) category, a task-irrelevant (distractor) category and a never task-relevant (neutral) category in fMRI (functional magnetic resonance imaging). Thus, neutral control baseline was operationalized by a category of visual objects that were not task relevant. Comparing target to neutral and distractor to neutral categories, they found target enhancement and distractor suppression.

1.2.5 Neural signatures of selective attention and neural speech tracking

Early studies found electrical signatures of selective attention in the human brain (Hillyard, Hink, Schwent, & Picton, 1973). Participants listened selectively to a series of tones and ignored concurrent tones in the other ear while their brain activity was measured. The specific neural activity arising from such acoustic stimulation is called an auditory evoked potential (AEP). AEPs are obtained by averaging multiple repetitions to the onset of the identical stimulus in the time domain of an electrophysiological signal and consist of a sequence of positive and negative deflections, called components. Each component can be seen as a representation of a neuronal activity along the auditory pathway (T. W. Picton, Hillyard, Krausz, & Galambos, 1974). Consequently, inferences about the underlying neuronal origins of the components are made possible by the components' latency. The cochlea, the auditory nerve, and the brainstem are connected to early (i.e., rapid) components within the first 10 ms. The auditory cortex responds to an auditory stimulus at latencies between 10 and 80 ms (i.e., midrange). Hillyard et al. (1973) found that selective attention modulates the component around 80 to 110 ms. The auditory cortex is the brain region most strongly linked to components after 80 ms, followed by the frontal and parietal brain regions (T. Picton et al., 1999). As a result, the processing of the signal in the auditory cortex is complex, which may cause components of the AEP to overlap in time. In auditory attentional research, the AEP is thus utilised to examine where, when and how attention affects the processing of auditory input. Based on these characteristics, the AEP could be described as an attentional filter. AEPs, with their characteristics of an attentional filter, could be used to investigate sub-mechanisms of selective attention such as the enhancement of targets or a suppression of distractors. However, their requirement for brief, isolated events is a significant obstacle to studying attentional systems in ecological contexts.

Research on the electrophysiology of attention to continuous signals such as speech has started in recent years thanks to advancements in computational processing (e.g., Crosse, Di Liberto, Bednar, & Lalor, 2016). Computational methods take advantage of the neural tracking phenomenon, which describes how electrical brain activity synchronises to specific features of sensory continuous stimuli (for review, see Obleser & Kayser, 2019). In auditory perception, neural tracking, or more precisely, speech tracking, refers to slow oscillations in the brain (delta 0.5–4 Hz and theta 4–8 Hz) that track the syllabic structure of speech (e.g., Ding & Simon, 2012; Lalor, Power, Reilly, & Foxe, 2009). In both, invasive (ECoG) and non-invasive electrophysiology (EEG/MEG) measures speech tracking can reliably be observed, since these methods have the temporal activity to reveal this neural activity (e.g., Ding & Simon, 2012; Golumbic et al., 2013). These low-frequencies (delta-theta oscillations) were observed not only in low-level auditory areas (for review, see Peelle & Davis, 2012) but also at higher processing, such as attentional control and language processing in inferior frontal cortex, anterior and inferior temporal cortex and inferior lobe (Golumbic et al., 2013).

It is still debated whether neural speech tracking arises from neural entrainment of neural oscillations or a superposition of responses evoked by sound features (Obleser & Kayser, 2019; Ding & Simon, 2014). A recent study supports the latter hypothesis (Zou et al., 2021). In this thesis, neural speech tracking refers to the mathematical approach that measures how well neural activity can be predicted from specific features of a speech stream. In this context, the measured electrical response and the amplitude envelope of the speech stream are linked via a linear filter, called the ”temporal response function” or TRF (Crosse et al., 2016). The TRF is frequently understood in close relation to accepted understandings of the auditory evoked brain response (described above) in traditional, event-related designs (Simon, Depireux, Klein, Fritz, & Shamma, 2007). By folding this temporal response function with the time series of the stimulus feature, a predicted EEG signal can be obtained. The correlation between the predicted and measured EEG signal is referred to as ”encoding accuracy” in a forward model, and it is interpreted as the strength of neural representation or neural tracking.

In the cocktail-party scenario, the listener can guide her attentional focus to select the talker of interest based on spectrotemporal features of speech (Shamma et al., 2011). The process of selective attention should therefore create a spectrotemporal filter to enhance the relevant audio stream by incorporating information from it (Lakatos et al., 2013). The TRF as filter should thus reflect both acoustic information and selective attention as a neural correlate of object formation and selection and represent speech as a whole auditory object (B. G. Shinn-Cunningham, 2008; Ding & Simon, 2012).

Both the anatomy and timing of neural activity reflect the acoustics and selection processes involved in the hierarchy of auditory processing (Nourski et al., 2014), as evidenced by the component structure of the temporal response function (TRF) which is similar to the auditory evoked potential (AEP). The early components of the TRF are associated with the core auditory cortex, while the later components correspond to activity in higher-order auditory areas. (Ding & Simon, 2012; Puvvada & Simon, 2017).

TRF and neural tracking of the attended stream is enhanced in multi-talker situations (e.g., Ding & Simon, 2012; Kerlin, Shahin, & Miller, 2010; Mesgarani & Chang, 2012). The attentional modulation is so powerful that even single trials are enough to decode to which speech stream listeners are attending (O'Sullivan et al., 2015). In addition, there is evidence that the degree of neural tracking correlates with speech intelligibility (Peelle, Gross, & Davis, 2013), with behavioural indices of speech comprehension (Etard & Reichenbach, 2019) and stronger speech tracking boosts trial-to-trial behavioural performance (Tune, Alavash, Fiedler, & Obleser, 2021).

While the attentional filtering mechanisms relating to the enhancement of target speech are relatively well understood, the mechanisms that enable ignoring (suppression of distractors) are much less understood. An implementation of a dual selective attention mechanism of target enhancement and distractor suppression could be beneficial at least for two reasons. First, any neural mechanism can only operate within finite limits. Hence, the rate at which a signal can be solely enhanced is limited by the upper bound. In contrast, a dual mechanism that enhances relevant signals (target) while suppressing irrelevant signals (distractor) can double the rate of the separation between the target and distractor by fully using the upper and lower limits. Second, a dual mechanism would be maximally effective over the entire dynamic range. For example, if the target and distractor were both at high levels of the dynamic range, further enhancing the target would be less successful than suppressing it, and if the target and distractor were both at low levels, further enhancing the target would be more effective than suppressing it (Tipper, Weaver, & Houghton, 1994).

In addition to the enhanced brain responses to target speech, responses to distracting speech were found to be sometimes suppressed (e.g., Rif, Hari, Hämäläinen, & Sams, 1991; Bidet-Caulet, Mikyska, & Knight, 2010). The EEG responses to target and distractor speech were found with opposite time-lags (Horton, D'Zmura, & Srinivasan, 2013). In combination with the findings from Lakatos et al. (2013) that the phase of slow neural ocillations alters the excitability of neurons, the opposite polarities of TRFs reflect enhancement of target and suppression of distractor speech (Horton et al., 2013).

Puvvada and Simon (2017) investigated cortical representations of multiple talkers in the auditory scene and found that the attentional background remains unsegregated. However, they did not manipulate the two streams in the background by task. In contrast, Kong, Mullangi, and Ding (2014) found evidence of suppression of responses to the unattended stream. By comparing unattended streams in a competing speech versus a passive listening condition, they found that the response to the unattended stream was attenuated. The effect of suppression is typically detected in the neural response around 200 msec after the start of the stimulus (Rif et al., 1991). Consistent with these findings, Fiedler, Wöstmann, Herbst, and Obleser (2019) demonstrated that varied signal-to-noise ratios result in different modulations of late TRF components, which are linked to cortical tracking of ignored speech. Suppressive mechanisms were also investigated in the auditory domain using alpha oscillations (neural oscillations 8-12 Hz; not phase-locked to the stimulus). The neural signature of alpha oscillations indicates the independent implementation of distractor suppression and target enhancement (e.g., Schneider, Herbst, Klatt, & Wöstmann, 2022; Wöstmann, Alavash, & Obleser, 2019).

1.3 Implications of presbycusis on the peripheral hearing and selective attention

Age-related hearing loss, also known as presbycusis, is a prevalent sensory impairment among older adults and a critical public health issue (Organization et al., 2017). Studies have shown that the prevalence of hearing loss greater than 35 dB HL is around 29% for women and 33% for men over 65 years of age (Homans et al., 2017). This hearing loss can lead to difficulties in understanding speech in everyday communication, which has been directly linked to presbycusis (Humes et al., 2012). Moreover, presbycusis is associated with long-term consequences such as social isolation (Weinstein, Sirow, & Moser, 2016), cognitive decline (Uchida et al., 2019), and depression (Lawrence et al., 2020). Of particular concern, hearing loss in midlife is the single largest modifiable risk factor for later dementia (Livingston et al., 2020). While hearing aids are the most common treatment for hearing loss (Dawes et al., 2015), recent advances in hearing aid technology have incorporated features such as dynamic compression and directional microphones to improve hearing in difficult situations (Dillon, 2008).

1.3.1 Object formation and selective attention in hearing impaired listeners

To this day, it remains unclear to what extent auditory perceptual and cognitive decline in hearing-impaired listeners causes difficulties in complex multi-talker situations and contributes to insufficient selective attention. Compared to normal hearing listeners, hearing-impaired listeners show reduced temporal and spectral sensitivity (Gaudrain, Grimault, Healy, & Béra, 2007; B. G. Shinn-Cunningham & Best, 2008). Broader frequency selectivity in hearing-impaired listeners leads to fewer independent channels, resulting in difficulties in segregating auditory objects (Pick, Evans, & Wilson, 1977; Rosen & Fourcin, 1986). In contrast, they have a comparable good ability to use temporal cues (Bacon & Gleitman, 1992; Turner, Souza, & Forget, 1995). Additionally, when the same envelope is used to modulate multiple spectral bands of competing sounds, hearing-impaired listeners experience increased perceptual interference (Hall III & Grose, 1994). In summary, mostly decreased spectral sensitivity in hearing-impaired listeners leads to impaired object formation, and they may compensate by relying more on temporal cues, even if their temporal processing is also decreased compared to normal hearing participants.

Given the close relationship between object formation and selective attention, it makes sense to assume that when object formation is compromised, so is selective attention. Unsuccessful object formation can lead to a decreased difference between targets and distractors. In this situation, attention is unable to act selectively and choose which object will be enhanced or suppressed (Desimone et al., 1995). Multiple studies have shown that hearing loss affects the ability to select target objects (for review, see B. G. Shinn-Cunningham & Best, 2008).

1.3.2 Changes along the auditory pathway due to presbycusis

The representation of modulation rates in the temporal envelope of non-speech sounds follows a hierarchical pattern along the auditory pathway, as indicated by evidence from studies on humans (Giraud et al., 2000) and animals (e.g. Schreiner & Urbas, 1986). Preferred amplitude modulation (AM) rate has been shown to decrease as one ascends the auditory pathway, with the olivary complex most responsive to faster modulation rates ($> 250Hz$), the inferior colliculus (IC) responding to ≈ 30–250 Hz, the auditory thalamus to AM rates at 16 Hz, and the primary auditory cortex tuned to low AM frequencies at ≈ 8 Hz (Giraud et al., 2000). This hierarchical organisation of the auditory pathway enables a decomposition of temporal modulations, with each level of processing acting as an AM filter (Dau, Kollmeier, & Kohlrausch, 1997a, 1997b). Due to age related hearing loss this processing is may weakened (Gaudrain et al., 2007; B. G. Shinn-Cunningham & Best, 2008). In the following, the most prominent changes due to presbycusis along the auditory pathway are explained.

When we are exposed to sound, acoustic signals reach our outer ear first and then move along the auditory pathway to the auditory cortex. The outer and middle ears are important to bundle and amplify sound on a mechanical level and are usually still preserved in old age (Rosowski, 1994). The amplified oscillations are transmitted via the footplate of the stapes into the cochlea and lead to the expansion of a wave therein. The wave leads to the activation of hair cells located at the basilar membrane. The basilar membrane shows a different stiffness based on its position within the cochlea. The stiffness is highest at the base (start) and lowest at the apex (end) of the basilar membrane. Directly related to the stiffness of the basilar membrane is its eigenfrequency. If the sound frequency transmitted from the outer world matches the eigenfrequency of the basilar membrane, the basilar membrane peaks and activates the inner hair cells at this location. This frequency-dependent spatial arrangement is referred to as tonotopy. Besides the inner hair cells (the actual sensory receptors), the outer hair cells are anatomically and functionally distinct hair cells that mechanically amplify low-level sounds (Kemp, 1986).

Outer hair cells lead to high sensitivity and sharp tuning of the travelling wave in a healthy cochlea, also called the active travelling wave. While the inner hair cells are usually well preserved at an advanced age, there is a causal link between people suffering from sensorineural hearing loss and the disruption of the mechano-electrical amplification process of the outer hair cells (Patuzzi, Yates, & Johnstone, 1989). This loss provides a cochlea-based explanation of two of the most common symptoms of sensorineurally impaired people. First, the absolute hearing threshold reached a higher level than in normal hearing. Second, the uncomfortable loudness level reached the same or even lower levels compared to normal hearing. The cause for this symptom is called ”loudness recruitment”. The loss of the sharp tuning of the active travelling wave leads to the activation of more inner hair cells (more inner hair cells are recruited). For hearing-impaired participants, this results in abnormal frequency resolution and increased loudness. Interestingly, the growth of loudness is similar (roughly linear) to normal hearing around the threshold (Buus & Florentine, 2002). In addition, cochlear compression is frequency-selective. The response to a tone presented at the center frequency (CF) is highly compressive, while a tone presented well below CF is roughly linearly represented in the basilar membrane response. Hence, the amount of compression applied to sounds depends on the relationship between the frequency of the sound and the CF (Oxenham & Bacon, 2003). In sum, for the sensorineural hearing-impaired, steeper growth of loudness and poorer frequency selectivity lead to an overall more linear response, while the healthy human cochlea is highly sensitive and non-linear to sound.

The most prominent age-related effect on hearing is the above described loss of outer hair cells. However, age-related decline affects all stages along the auditory pathway. Cochlear synapses between inner hair cells and the cochlear nerve decline with age, which leads to degraded temporal envelope cues early in the auditory pathway (Parthasarathy, Bartlett, & Kujawa, 2019). Past the cochlea, animal and human data reveal age-related changes at the level of the cochlear nerve, medulla, pons, midbrain, and inferior colliculus (Peelle & Wingfield, 2016). The envelope following response (EFR) as part of the auditory brain stem response (ABR) is an evoked potential produced by periodic or quasi periodic stimuli and commonly used to evaluate the auditory periphery. Ageing also affects the EFR and ABR responses. Neural phase-locking to the envelope and temporal fine-structure are impaired in people with mild to moderate presbyacusis at the level of EFR (Ananthakrishnan, Krishnan, & Bartlett, 2016).

The primary auditory cortex receives inputs from the ascending auditory pathway and transmits them to cortical processing. Auditory information only reaches perception if it is processed by the cortical area. Hence, changes in the auditory cortex due to aging and hearing loss are of particular significance for auditory perception. The processing strategies within the auditory cortex are complex and still not fully understood (e.g., Bizley & Cohen, 2013; Schreiner, Read, & Sutter, 2000). Ageing and hearing loss affect the auditory cortex in multiple ways. For instance, ageing is associated with changes in evoked responses and affects the structure, tuning, selectivity, and temporal processing of the auditory cortex (for review, see Peelle & Wingfield, 2016). The effects of ageing and hearing loss on the auditory cortex are multifaceted and complex, in this thesis, I will focus on how they affect neural speech tracking, which is covered in the section below.

1.3.3 Neural speech tracking in hearing impaired listeners

The growing body of research on how age-related hearing loss affects the neural tracking of speech in the auditory cortex has not been fully answered yet. An earlier study reported larger neural tracking of the ignored stream for hearing impaired listeners, which results in smaller differential tracking between the attended and ignored streams (Petersen, Wöstmann, Obleser, & Lunner, 2017). Recent research has found that hearing-impaired listeners have larger neural tracking responses to target speech (Decruy, Vanthornhout, & Francart, 2019; Fuglsang, Märcher-Rørsted, Dau, & Hjortkjær, 2020). Comparing hearing impaired participants with age-matched normal-hearing peers, hearing-impaired listeners had increased neural tracking, delayed neural responses to continuous speech in quiet and the latency also increased with the degree of hearing loss (Gillis, Decruy, Vanthornhout, & Francart, 2022). In additon, Schmitt, Meyer, and Giroud (2022) reported enhanced speech tracking with increasing hearing loss and suggested that the hearing impaired rely more on the tracking of slow modulations in the speech signal to compensate for their hearing deficit. Additionally, brain tracking of the fundamental frequency of the voice is linked to cortical compensation for hearing loss but not to age (Van Canneyt, Wouters, & Francart, 2021). On the other hand, it was also demonstrated that linguistic and acoustic speech representational neural tracking declines with age (Gillis, Kries, Vandermosten, & Francart, 2023). Tune et al. (2021) reported no increased neural tracking with age or hearing loss. Interestingly, Decruy et al. (2019) and Gillis et al. (2023) used the same data set but found contradictory results depending on the modelling approach used to obtain neural tracking. Therefore, the opposite effects could result from the method selection, but also, for instance, various experimental design decisions could have an impact on the outcomes. Even if the current state of this collection of research does not clearly indicate what occurs to the brain's ability to track speech, it does show that the ability is not lost due to hearing loss. Further, hearing aid algorithms also affect neural speech tracking. Petersen (2022) found that hearing aid directionality improves neural speech tracking in hearing impaired listeners.

1.3.4 Hearing aid algorithm

Modern hearing aids do not simply enhance the sound from the outer world but use a wide range of sophisticated signal processing algorithms in combination with multi-microphone technology to deliver the best possible speech comprehension and hearing quality to meet the individual needs of hearing-impaired people.

Multimicrophone algorithms, for example, can use the spatial information of the sound scene in addition to the spectrotemporal information. The additional use of spatial information is especially beneficial for performance when target sound and distractor sound are locally separated (e.g., Jensen, Høydal, Branda, & Weber, 2021).

Another widely used hearing aid algorithm is dynamic range compression (DRC). DRC is an audio signal processing algorithm that amplifies quiet sounds while reducing the intensity of loud sounds. Despite having higher hearing thresholds, hearing impaired listeners with presbycusis perceive sounds as more intense and louder than healthy hearing listeners (loudness recruitment, see section 1.4.2). In hearing impaired listeners suffering from presbycusis, the combination of increased hearing thresholds and loudness recruitment results in a decreased dynamic range. DRC compensates for a decrease in dynamic range by applying a lower gain to sounds with a higher intensity (Kates, 2005). However, DRC also leads to unwanted side effects. The compression of the sound signal is not instantaneous. Attack and release times lead to abrupt changes in the onset and offset of the signal. In addition, compression directly affects the envelope of a speech signal, since it reduces the amplitude modulation depth of speech (Stone & Moore, 1992). A change of the speech envelope could impair the processing of the speech stimulus, since the speech envelope is associated with speech comprehension and is not only an inoperable acoustic feature of the signal (for review, see Poeppel & Assaneo, 2020).

1.4 Research questions

The two thematic foundations that form the basis of this thesis' research questions are: First, investigate the mechanisms of selective attention utilising behavioural responses and electrophysiological recordings in a novel psychophysically augmented continuous speech paradigm. Second, this thesis aims to explore the neural mechanisms of selective attention to speech when speech streams are degraded by a compression of their temporal envelope. In total, 73 participants took part in this thesis.

This thesis' first section aims to provide an answer to the following: How is selective attention implemented in a multi-talker situation? Study 1 looks into whether the behavior-supported neural tracking signatures of selective attention represent target enhancement, capture, and then suppression of the distractor, or both. A neutral control baseline in the form of a never-task relevant stream was operationalized to distinguish between the two potential sub-mechanisms of target enhancement and distractor suppression. To obtain finely-resolved behaviour to continuous speech, short repeats were embedded in the speech stream, which the participants had to detect.

The second part of the thesis tries to shed light on the following question: How does amplitude compression of speech impact the neural and behavioural signatures of selective attention in a situation with multiple talkers? To answer this complex question, this study is divided into several sections. In the first section, it was tested whether different amplitude compression ratios affected neural tracking. In the next section, the interplay between selective attention and amplitude compression was investigated in normal hearing listeners. Using a computational model of the human auditory periphery, the neural fate of loudness matched amplitude compressed speech was simulated. Of particular interest, the relationship between selective attention and amplitude compression was investigated in hearing impaired listeners. Behavioural responses and neural tracking in younger, normal hearing and older, hearing impaired listeners were compared to test whether the manipulation of compression has a different impact on listeners suffering from presbycusis. Finally, in an online study, the loudness matching algorithm used for amplitude compressed speech was behaviorally evaluated.

2 General methods

2.1 Psychophysically augmented continuous speech stimuli

Speech is the most important medium in human communication, which makes it one of the most salient and behaviorally relevant signals in human environments. For this reason, a fundamental goal in neuroscience is to investigate how cognitive processes deal with naturalistic speech. Hamilton and Huth (2020) propose that the use of natural stimuli will be the future of auditory neuroscience. In accordance with this, speech in the form of continuous narrative stories was presented to participants in this thesis. This was mainly done for two reasons. First, this work should contribute to aiding hearing impaired listeners and addressing their challenges in real-life listening scenarios. Second, advances in computational modelling enable the analysis of neural responses to continuous speech.

People are exposed to narrative stories outside of the experimental context, which makes narrative stories more ecologically valid compared to isolated words or isolated sentences. Listening scenarios that are considered naturalistic in an ecologically valid sense, on the other hand, include social interactions and unpredictable changes in conversational partners. Hence, narrative stories are a medium on which we operate and are placed on a continuum between discrete and naturalistic stimuli. We would argue, however, that narrative stories are more on the naturalistic spectrum due to their continuous character and real-life appearance.

In comparison to trial-based designs, the presentation of continuous speech improves ecological validity. Contrarily, a significant drawback of continuous speech paradigms up to this point has been their typical lack of rich behavioural data (Hamilton & Huth, 2020). Commonly, comprehension questions about the speech stream's content are asked infrequently or later. This makes it difficult to determine whether or not the neural responses are relevant to the task at hand. On the one hand, there are the fine-sampled neural recordings, and on the other, there are the very discrete comprehension questions. This presents a particularly difficult problem for the study of brain-behavior relationships. Within this thesis, I have attempted to address this challenge by including short, repeated segments of speech into the speech streams inspired by Marinato and Baldauf (2019). Participants had to detect the embedded repeats as fast as possible in the target stream and had to ignore them in the streams in the attentional background. As a result, we were able to measure the response times and hit and false alarm rates for the repeats embedded in different speech streams.

In the first experiment, we presented three narrative stories simultaneously and spatially separated (-45, 0, 45) to the participants in the free field. Narrative stories were spoken by different untrained male speakers. Participants could therefore rely on spatial and spectral cues to segregate the speech streams. We aimed to investigate the mechanisms of selective attention by implementing a neutral baseline. To separate the neutral and distractor streams on the behavioural level, we were reliant on false alarms in participants' behaviour evoked by these streams. To achieve this, we needed a comparable level of stream segregation. As a result, we chose a comparable small spectral and spatial separation. In the second experiment, we presented two narrative stories concurrently and spatially separated (0, 180) to the participants in the free field. We presented narrative stories spoken by professional male and female speakers. Participants could therefore rely more heavily on spectral cues to segregate the speech streams. This was done mainly for two reasons. First, hearing-impaired participants were to be measured in this setup. In complex listening situations, hearing impaired participants typically have difficulty listening and attending to speech. Second, hearing aids can more effectively apply separate hearing aid algorithms in the front versus the back semi-field.

2.2 Electroencephalograpy

Temporal properties of human speech are important characteristics for the human brain in speech perception. Slow envelope fluctuations were shown to contribute significantly to the perception and attention of a human speech signal (e.g., Rosen, 1992; Hickok & Poeppel, 2007; Abrams, Nicol, Zecker, & Kraus, 2008; Giraud & Poeppel, 2012). Investigating, the neural processing of the speech envelope requires a brain imaging technique with a fast temporal resolution. In this thesis, electroencephalography (EEG) was used as the method of choice since EEG is considered to have excellent time resolution.

Hans Berger is the developer and eponym of the EEG. He measured the first human EEG back in 1924 (H. Berger, 1930). Since then, the technical basis has not changed much. Voltage fluctuation is measured at least between two electrodes. The magnitude of the signal measured on the scalp ranges approximately between 5 and 100 μV. EEG recording systems use differential amplifiers to reduce the impact of noise from the ground circuit. To achieve this, a reference electrode is used in addition to the active electrodes. The reference electrode also records the signal relative to the ground electrode and therefore also contains noise from the ground circuit. By subtracting the signal from the reference electrode from the signals from the active electrodes, the noise is cancelled out, since it is approximately the same in both the active and reference electrodes.

Changes in the postsynaptic potentials of cortical pyramidal cells are the source of the EEG signal. Primarily, pyramid cells with a vertical orientation produce vertical dipoles. Only a sum of these vertically oriented dipoles is relevant for the EEG. This means that the EEG is not measuring primary currents (action potentials) but rather secondary currents (postsynaptic potentials). In contrast to other methods in neuroscience such as fMRI, EEG has a high temporal resolution because it measures the electrophysiological process directly (Gevins, Leong, Smith, Le, & Du, 1995). However, the spatial resolution is limited due to the further spread caused by the limited conductivity of the skull and scalp (Ahlfors, Han, Belliveau, & Hämäläinen, 2010).

For EEG recordings in this thesis, the SMARTING amplifier (mBrainTrain, Belgrade, Serbia) was connected to 24 electrodes of the EEG-cap (Easycap, Herrsching, Germany; Ag-AgCl electrodes placed according to the 10-20 International System). This is a mobile EEG-system that transfers the signal via Bluetooth to a recording computer. This feature gives the system a relatively high degree of flexibility in terms of application. In this thesis, we were mostly interested in answering ”how” questions and not ”where” questions (Bizley & Cohen, 2013). In additon, we mainly used encoding models, which estimated a model for each single channel (see below). As a result, the relatively low number of electrodes posed no significant restriction for us.

2.2.1 EEG data preprocessing

The human brain is not a linear, time-invariant system that responds consistently to all input at all times. On the contrary, the human brain is quite nonlinear and time-variant. This and external noise pose a challenge analysing EEG recordings. The aim of the preprocessing is to get rid of noise from irrelevant brain areas as well as from non-brain sources such as muscle activity or external noise such as electromagnetic interference. There is no general optimal choice for the different pre-processing steps. The preprocessing has to match the research question and underlying assumptions and can thus vary between experimental designs. We used high- and low-pass Hamming-window FIR-filters with cutoff frequencies of 1 and 40 Hz, respectively, in this thesis (Fiedler et al., 2019). Further, recordings were re-referenced to mastoid electrodes to prepare for independent component analysis (ICA), which is an algorithm to split up a mixture of multiple sources of variance into components (Comon, 1994). The data were then divided into 10-second epochs. Noisy epochs were excluded for ICA by visual inspection. After performing an ICA on the remaining data, components identified as artifactual by visual inspection were removed. In order to prepare for linear modelling, a 1-Hz-highpass and 10-Hz-lowpass filter were then applied. High-pass filtering eliminates drifts, whereas low-pass filtering removes high-frequent noise. Since previous research has shown that neural activity phase-locked to the speech envelope typically occurs below 10 Hz, we here chose a 10-Hz-lowpass cut-off(e.g., Golumbic et al., 2013; Simon et al., 2007).

2.3 Feature extraction from continuous speech

In terms of physics, sound is a pressure change that travels through a transmission medium like a gas, liquid, or solid as an acoustic wave. These sound waves are characterised by frequency, amplitude, speed, and direction. The smooth curve that delineates the extremes of an oscillating sound signal is called its envelope. Of course, these characterizations also apply to speech. Along the auditory pathway from the outer ear to the auditory cortex, these physical characteristics of speech are transformed, grouped, selected, and turned into meaning.

In linguistics, a syllable is an organisational unit for speech sounds. They are also referred to as ”building blocks” of speech. The speech envelope captures this syllable structure of speech. Although the process of creating a complete model of the auditory pathway leading to the auditory cortex is still in its early stages (Verhulst, Altoe, & Vasilkov, 2018), the speech envelope is associated with phase-locked neural activity (e.g., Golumbic et al., 2013). However, the precise method for obtaining a speech signal’s envelope is far from standardised and is partly up to the analyst’s judgment.

A straightforward approximation to the broad-band temporal envelope of speech is the magnitude of the analytic signal. Here, we used the more sophisticated approach of utilising an auditory model of the cochlea that extracts a cochleogram of sound (summed to a broad-band envelope), which was shown to improve the representation of the speech envelope in the brain (Chi, Ru, & Shamma, 2005; Biesmans, Das, Francart, & Bertrand, 2016). It has been demonstrated that the auditory cortex is extremely responsive to the rate of change of the envelope (Howard & Poeppel, 2010, 2012) and that onsets produce the strongest neural responses and TRF components with the highest similarity to a classical ERP (Fiedler et al., 2019; Chalas et al., 2023). Hence, in this thesis, we focused on the envelope onsets, which are represented as the first half-wave rectified derivative. More complex or abstract speech features can be modelled in addition to low level acoustic features like the onset speech envelope. Stick-functions, for example, can be used to model higher-order linguistic features such as phonemic or semantic features based on their occurrence in the continuous speech signal (Brodbeck, Hong, & Simon, 2018).

2.4 Neural response to continuous stimuli

The world we live in is continuous, and as such, the human experience unfolds continuously in time. During speech perception, the human brain transforms the continuously varying speech signal into meaning. The brain response thus represents a continuous function of the input stimulus. Nevertheless, brain responses are not a direct representation of the stimulus but reflect multiple transformations of that stimulus. Although not all cortical neurons fit this description, there are neurons in the cerebral cortex's sensory regions that act as sensory transducers that are activated by sensory stimulation. The simplest transducers are linear, and a linear time-invariant system is characterised by its impulse response (Ringach & Shapley, 2004). Importantly, the impulse response provides a full characterization of a linear time-invariant system. The impulse response can be measured in many different ways (Figure 1). In the following, two methods commonly used in neuroscience are described. First, it can be measured as the response to a brief stimulus. In auditory electrophysiology, this approach is basically used to measure ERPs in the event related framework (Hillyard et al., 1973). Second, the impulse response can be measured by cross-correlating a broadband white noise with it corresponding output. In principal this approach is associated with the TRF or encoding/decoding framework (Crosse et al., 2016), which is described in more detail in the following.

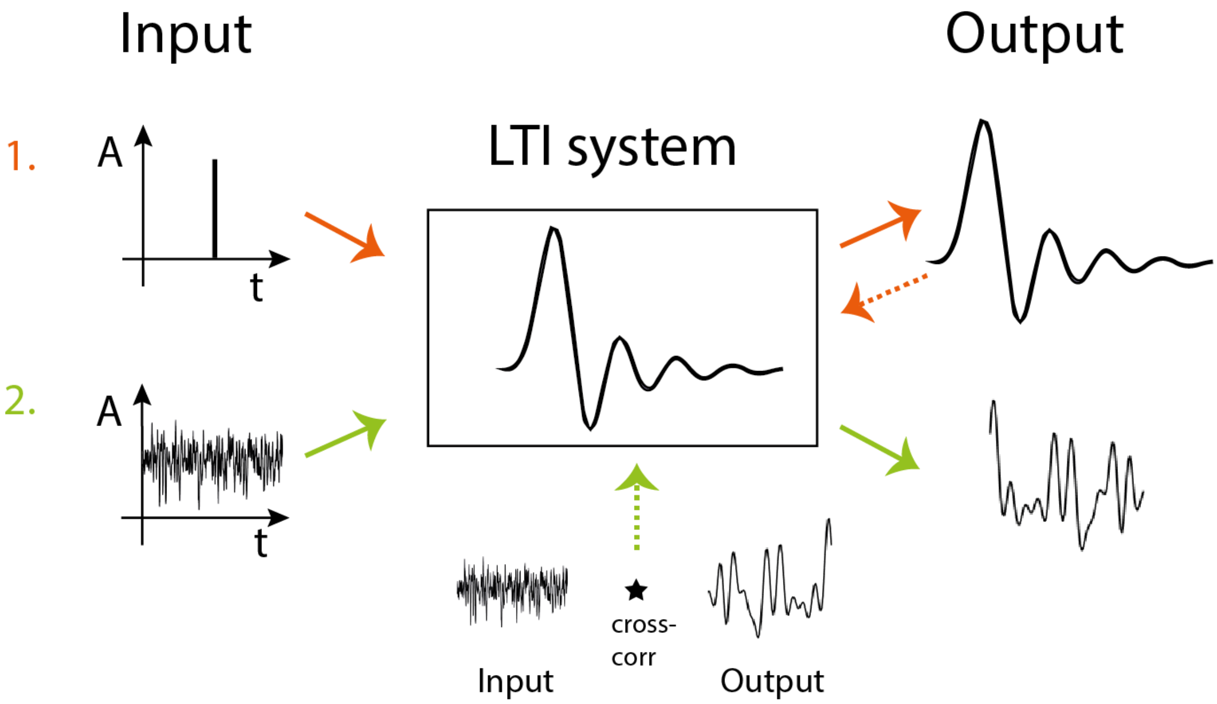


Figure 1: Measurement of the impule resonse of a LTI System

Two ways to measure the impulse response. A Dirac impulse is used as the LTI system's input in the first pathway (red), which shows how the impulse response is measured. A theoretical signal known as the Dirac impulse has an infinitely short period of time, an infinitely large amplitude, and an infinitely continuous frequency. It is the convolution's invariant element. The impulse response of the LTI-system itself is the output if the system is now simulated with a Dirac impulse (input). White noise is used as input to obtain the impulse response in the second pathway (green). The resulting output of the LTI-system is a superposition of the input signal and the impulse response function. A cross-correlation between the input and output results in the impulse response function of the LTI-system.

Continuous stimuli, such as continuous speech, are a challenge for event-related analysis frameworks but can be analysed in the framework of encoding/decoding models. Prior to such modelling, a critical choice between two options—even though they both fall under the same machinery—must be made.

2.4.1 Encoding versus decoding models

To estimate the neural response to continuous stimuli, there are two complementary approaches established in the field of auditory neuroscience (Crosse et al., 2016). The two models differ in their direction of stimulus-response mapping. The decision to use one of the two models depends on the question asked. Are we interested in how well the neural response can be predicted from multiple stimulus features? This approach refers to encoding (or forward) models. It is also called forward model since it reflects the stimulus-response mapping from stimulus to neural response. Or, are we interested in how well the stimuli feature can be reconstructed from neural response? This approach maps the neural response to the past stimulus (opposite direction to encoding model) and is therefore fittingly called the decoding (or backward) model (Figure 2).

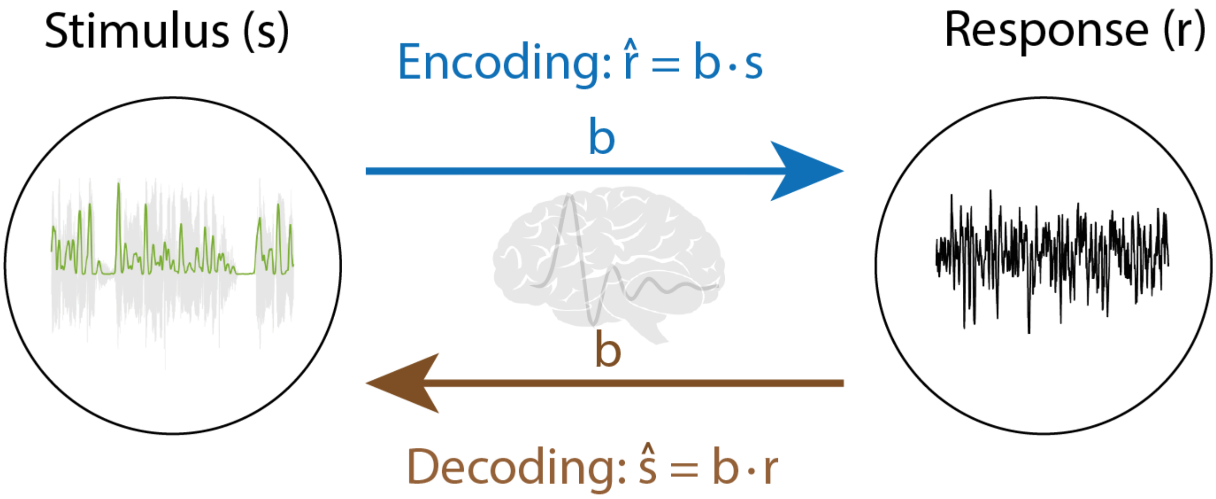

Stimulus (s)
Response (r)
Encoding: $\hat{r} = b \cdot s$
b
b
Decoding: $\hat{s} = b \cdot r$

Figure 2: Encoding decoding approach

Encoding versus decoding approach. Encoding models (”write-in”) work in the direction of information flow when studying sensory systems. What kernel/impulse response gets me from a given stimulus (s) to a neural response (r)? Decoding models (”read-out”) predict stimulus features using patterns of brain activity. How can I decipher from a given neural response (r) which stimulus (s) was present? In the regression approach, we are aiming for the linear operator (b), which gets us from stimulus (s) to a estimated brain response ($\hat{r}$;encoding) or from a recorded brain response to a predicted stimulus ($\hat{s}$;decoding). Here, b is estimated via a ridge regression approach.

Both approaches are mathematically similar and aim for a kernel between stimulus and neural response. However, both approaches have their own advantages, and the decision to choose one over the other depends not only on the asked question but also on other more detailed considerations derived from the properties of each model. The temporal response function reflects the encoding model in the mTRF-regression-framework (Crosse et al., 2016). They define how a brain response changes with each one-unit change in a specific stimulus component, and they are evaluated independently for each EEG channel. Their beta weights provide an intuitive, neurophysiological interpretation that is comparable to an event-related potential (Simon et al., 2007). This enables us to directly compare TRFs between conditions and to separate brain responses to even concurrently presented stimuli. This univariate prediction at each EEG channel also allows us to interpret the topography of the encoding brain. Further, correlations between stimulus features are explicitly accounted for in the encoding model (for review, see Holdgraf et al., 2017). On the other hand, decoding models take correlations between EEG channels into account, since it maps the data from all EEG simultaneously. This makes the pre-selection of certain channels redundant. The multivariate characteristic of decoding models leads to increased sensitivity since channels are weighted based on their relevance (Pasley et al., 2012). Decoding model weights are not directly interpretable in terms of a studied brain process. However, it is possible to transform decoding models into encoding models to facilitate their interpretation in terms of a brain process (Haufe et al., 2014).

Within this thesis, mostly encoding models were used mainly for two reasons. First, we were interested in studying the effects of attentional and amplitude compression manipulations on the morphology of the TRFs. Second, our repeat detection task led to potential ERPs evoked by the repeats and to motor activity by the button presses to detect repeats. The encoding model approach takes correlations with this potential confound into account.

2.4.2 Temporal response function

The kernel, or impulse response function, that characterises the influence of each unit in the predictor on the neural response is called TRF. The TRF is usually unknown in contrast to the features of the continuous stimulus and the neural response. The TRF algorithm exploits this to estimate the TRF that shows the best prediction of the neural response from the predictor variables. The central assumption of the TRF approach is that a convolution of the predictor variable (stimulus representation), s, with the TRF (kernel), β, and a residual response ϵ results in the dependent variable (neural response), r.

$$r = \beta * s + \epsilon \tag{1}$$

In discrete time, the encoding model is represented as:

$$r(t,n) = \sum_{\tau} \beta(\tau,n)s(t-\tau) + \epsilon(t,n) \tag{2}$$

where $r(t,n)$ is the neural response, sampled at times $t = 1...T$ and at channel n. The TRF $\beta(\tau,n)$ describes the linear transformation of the ongoing stimulus ot the ongoing response for a given range of time lags τ, relative to the transient occurrence of the stimulus feature $s(t)$. The residual response at each channel that the model is unable to account for is denoted by the error term $\epsilon(t,n)$.

The TRF (β) is estimated by defining a measure of the error term. The mean-squared error (MSE) is here the target to minimize the error between the actual response $r(t,n)$ and the response predicted by the convolution $\hat{r}(t,n)$.

$$MSE = min\,\epsilon(t,n) = \sum_{\tau}[r(t,n) - \hat{r}(t,n)]^2 \tag{3}$$

This method is solved using reverse correlation (De Boer & Kuyper, 1968). As a result, the TRF is calculated using the subsequent matrix operation:

$$\boldsymbol{\beta = (S^T S)^{-1} S^T r} \tag{4}$$

where $\boldsymbol{\beta}$ is the TRF $\tau_{window} \times N$ matrix with each column containing a TRF for each channel. $\boldsymbol{S}$ is a $T \times \tau_{window}$ matrix containing the stimulus representation with time-lagged repetitions. $\boldsymbol{r}$ is a $T \times N$ matrix containing the column-wise arranged neural data.

2.4.3 Regularization

Only very broadly defined terms have been used up to this point to refer to fitting linear models that map between presented stimuli and measured neural responses. A comparatively large number of regressors is modelled as a result of the inclusion of various time lags. These regressors may also be highly correlated because many stimulus regressors, including the acoustic envelope (a non-white stimulus), exhibit significant auto-correlation and neighbouring EEG channels pick up similar signals. As a consequence, a simple cross-correlation would result in a temporal smearing of the TRF. The solution is to divide out the covariance structure of the stimulus. Therefore, the method referred to ridge regression is used to reliably estimate the TRF (Crosse et al., 2016).

Regularization introduced an additional term into equation 4 to solve the ill-posed estimation problem and prevent overfitting. Inverting the auto-covariance matrix $\boldsymbol{S^T S}$ is numerical instable. To take care numerically, this means adding a smoothing term to reduce variance in the estimate. In addition, this smoothing term penalises large differences between neighbouring TRF values, which makes the TRF less specific and easier to generalise (Hastie, Tibshirani, Friedman, & Friedman, 2009). Regularization can be introduced by adding the smoothing term, as shown below:

$$\boldsymbol{\beta = (S^T S + \lambda\, I)^{-1} S^T r} \tag{5}$$

$\boldsymbol{I}$ is the identity matrix and λ is the regularization parameter. λ can have values that fall within the range of $[0; \infty]$. A $\lambda = 0$ would result in a zero matrix for the identity matrix and since zero is the neutral element in addition $\lambda = 0$ would have no effect on $\boldsymbol{S^T S}$. In other words, the effect of $\lambda = 0$ would be equal to ordinary least square regression as in eq. 4. On the other hand, $\lambda > 0$ would increase the regularization. Interestingly, regularization appears to have a greater effect on the decoding model (Wong et al., 2018).

Practically, iterative model training and testing for a given set of lambda values empirically determines the ideal level of regularization. As an alternative, one could decide to alter the range of tested values in accordance with the regressors' autocovariance structure (Fiedler et al., 2019).

2.4.4 Training and testing

To test model performance, model training is complemented with model testing on held-out data. In order to prevent overfitting, model training aims to generalise the model to new data. In model testing, the trained encoding models are convolved with a unseen stimulus segment to predict a EEG response per channel. Cross-validation is a popular method for accomplishing this. (Varoquaux et al., 2017). Model performance is evaluated by two different validation metrics: Pearson's correlation and mean squared error (see above). In the TRF-framework the mean squared error is usually used to find the optimal ridge parameter λ and Pearson's correlation of model predictions with the EEG-signal to obtain the correlation-based measure of neural tracking.

2.4.5 Neural tracking

In the literature, neural tracking refers to both the temporal response function and the correlation-based accuracy as strengths of neural representation (e.g., Obleser & Kayser, 2019; Tune et al., 2021; Brodbeck et al., 2018). Nevertheless, in this thesis, neural tracking refers rather to correlation-based accuracy, which is of course closely linked to the TRF. The predicted EEG signal computation is expressed as a matrix operation as follows:

$$\hat{\boldsymbol{r}} = \boldsymbol{S}\boldsymbol{\beta} \tag{6}$$

where $\hat{\boldsymbol{r}}$ is the predicted EEG response, $\boldsymbol{S}$ the stimulus matrix (e.g, time-lagged versions of broadband envelope) and $\boldsymbol{\beta}$ the TRF matrix.

As described above the Pearson's correlation is used as a measure for neural tracking by quantifying the relationship between predicted ($\hat{r}$) and measured EEG signal (r). Mathematically, the Pearson's correlation coefficient is denoted as:

$$r_{pearson} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

where $x = \hat{r}$, $y = r$ and n the number of samples.

The Pearson's correlation coefficient ($r_{pearson}$) reflects the correlation-based accuracy and, as such, neural tracking. Importantly, negative coefficients do not mean the same thing as positive coefficients, in contrast to the most typical case of Pearson correlation. Negative coefficients would indicate a different polarity of the predicted and measured EEG signal, which can only explained as noise. However, for positive correlations, the greater the neural tracking, the higher the signal-to-noise ratio in the recorded EEG. Typically, the correlations range around 0.05 in the encoding approach (e.g., Jessen, Obleser, & Tune, 2021; Fiedler et al., 2019).

2.5 Dynamic range compression

Dynamic range compression is an audio signal processing technique that amplifies quiet sounds while reducing the intensity of loud sounds, thereby decreasing the dynamic range of an audio signal. To adjust dynamic range compression algorithms, a number of parameters can be adjusted (Kates, 2005). In the following, the most central parameters (threshold, ratio, and attack and release times) of a digital compressor are described.

The threshold is the point at which the compressor starts working. The threshold is set in dBFS for digital compressors. A lower threshold (e.g., -50 dBFs) means that the signal is more affected by the compression than a signal at a higher threshold (-10 dBFs). In other words, the higher the threshold, the less the compression on the signal (Figure 3 A).

A compressor reduced the gain by a certain ratio. A ratio of 2:1 means that if the input level is 2 dB over the threshold, the output signal is reduced to 1 dB over the threshold. In other words, 2:1 could also be stated as 1 over 2 which means that the output level is reduced by 50% over the threshold. A ratio of ∞:1 is often referred to as a limiter that reduced each input signal above the threshold almost to threshold level (Figure 3 A) .

The compressor does not typically react instantaneously to a given input. Attack and release times determine how fast a compressor acts. The attack time is the period that the compressor needs after the threshold is reached to reduce the gain according to the compression ratio. The release time determines how long it takes for the compressed gain to return to linear gain after the input level is reduced (Figure 3 B).

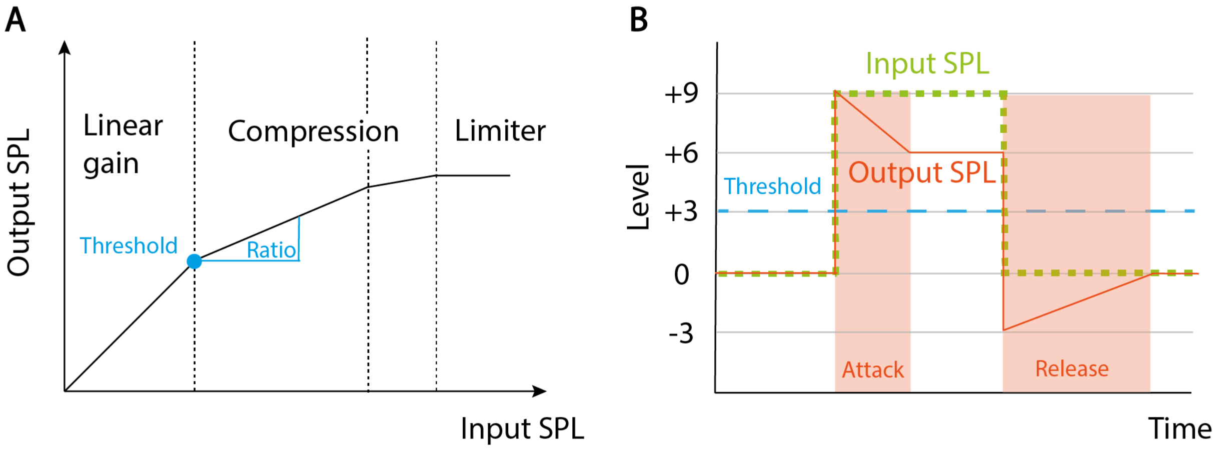


Figure 3: Dynamic range compressor

A. Input/Output curve hearing aid compressor amplifier. The curve of an hearing aid compressor usually contained at least three areas. First, low level inputs are amplified linear. Second moderate to high inputs are compressed. Third, very loud signals are cut-off via a limiter. Lower threshold (or knee) determines operating point of the compressor. The slope of the compressor determines compression ratio. Upper threshold determines operating point of the limiter. **B. Attack and release compressor.** Attack and relase periods are illustrated over time. Threshold indicates operating point of the compressor. The compressor do not apply the full compression to the signal that exceed the threshold. Compressor gradually reduced amplification over until the ratio is reached. This period is called attack time. Also, the compressor do not stop instantly compressing the signal when the input falls below threshold. This period is called release time.

There is no standard for how these parameters have to be set. It depends on their application. Compression is commonly used in sound recordings, instrumental amplifiers, broadcasting, and hearing aids. Here, we focused on the application of digital amplitude compression to hearing aids.

In hearing aids, a compressor is usually used to compensate for loudness recruitment by bringing back the levels into the listener's hearing range. Applying linear gain only to compensate for hearing loss would involve a compromise between audibility and comfort. Low-level audibility would be sacrificed to avoid loudness discomfort for high-level sounds. Hence, most modern hearing aids use a combination of linear gain and amplitude compression since it enables the audibility of low level sounds and the reduction of high-intensity sounds (Figure 3). Up to 20 or even more channels are employed in modern hearing aids. Usually each channel can apply frequency-dependent compression (e.g., Kollmeier, Peissig, & Hohmann, 1993). However, amplitude compression also has undesired side effects such as reduction of amplitude modulation depth, distortion of the envelope shape and abrupt changes in the onsets (overshoot) and offsets (undershoot) of sounds (Stone & Moore, 1992).

The processing power of hearing aids has increased over the years. Beamforming approaches enable modern hearing aids to apply spatial-dependent signal processing such as compression on different positions. For instance, beamformer processing technology divides incoming acoustic signals into two distinct streams. The sounds in one stream are primarily coming from the front of the wearer, while the sounds in the other stream are coming from the back. A dedicated processor is used for each stream to examine the characteristics of sound coming from every angle (Jensen et al., 2021).

Here, we were interested in combining the use of beamformer processing technology with the application of various compression ratios to speech streams at distinct locations. We experimented with simulating such processing over free-field loudspeakers.

2.6 Peripheral auditory modelling

There are many different types of models for the human auditory periphery, ranging from very simple functional descriptions of auditory filtering to intricate computational models of cochlear mechanics, inner-hair cell (IHC), auditory nerve (AN), and brainstem signal processing. To investigate the peripheral fate of unprocessed and compressed signals in normal hearing and hearing impaired, we used computational modelling of the human auditory periphery (Verhulst et al., 2018).

The model of Verhulst and colleagues is a complex model that enabled us to model different stages along the auditory pathway, and importantly, hearing loss. The model outputs are: single-unit simulations of the auditory nerve (AN), cochlear nucleus (CN), inferior colliculus (IC) and envelope following responses (EFR). Here, we are especially interested in the output of the EFR. The processing pipeline (Figure 4 A.) is summarised in the paragraphs that follow.

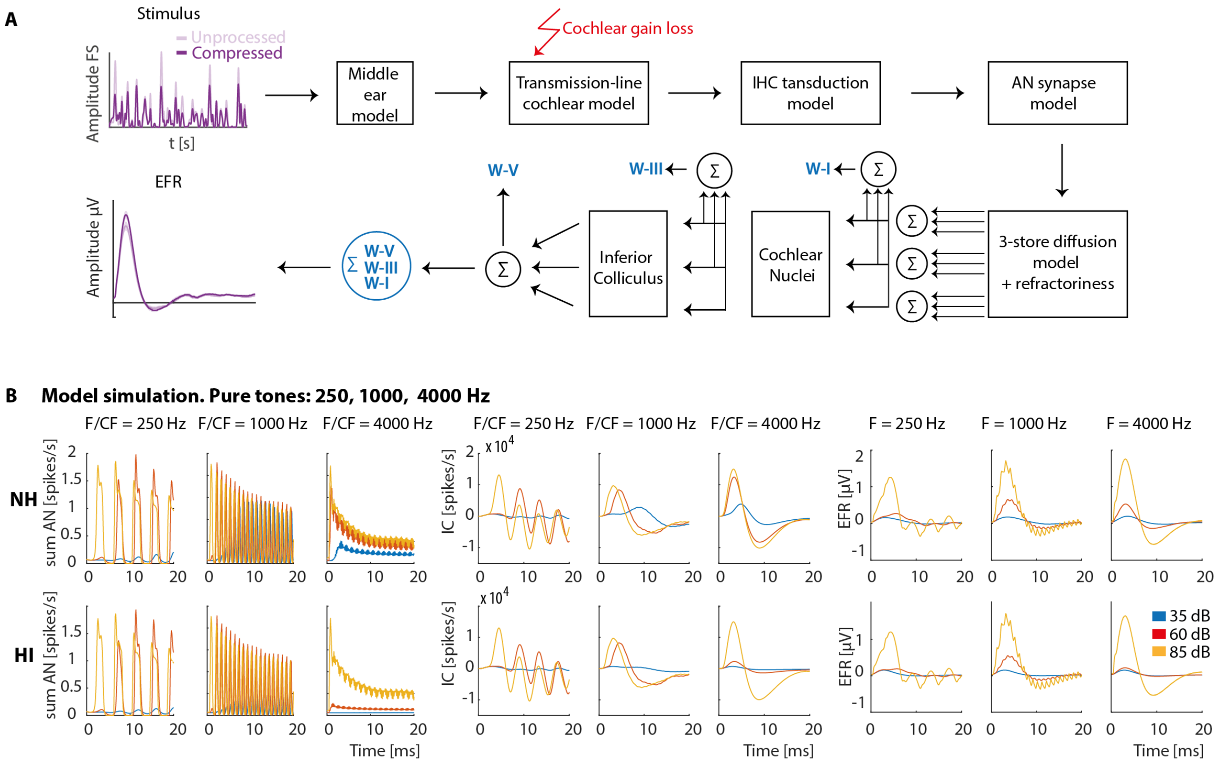


Figure 4: Human auditory periphery model Verhulst

Human auditory periphery model. With amplitude compressed and uncompressed speech as inputs and the EFR in the time domain as the model output, Figure **A.** depicts the various stations of the human auditory periphery model in a simplified manner (for details, see Verhulst et al., 2018).**B.** For the given input of pure tones (250, 1000, and 4000 Hz) at three sound pressure levels (35 (blue), 60 (red), 85 dB (yellow)), we simulated model outputs (AN, IC, and EFR). This was simulated for normal hearing (NH) and hearing loss (HI). The hearing loss was set in accordance to a typical mild-to-moderate sloping hearing loss starting from 1 kHz to 8 kHz 35 dB HL.

The stimulus first passes through a first order middle-ear bandpass filter before entering a transmission-line cochlear model with cochlear compression and tuning estimates based on human otoacoustic emissions (OAEs). The transmission-line model simulates waveforms across 1000 discretized cochlear sections that span the human range of hearing (Greenwood, 1990) as well as OAEs that can be compared to human data at the middle-ear filter output. The IHC-AN synaptic complex model, which includes a biophysical description of the IHC membrane potential, a synaptic exocytosis (models acitve transport of neurotransmitters) model (Beutner, Voets, Neher, & Moser, 2001), as well as a three-store diffusion (Meddis, 1986) and refractoriness model (Peterson, Irvine, & Heil, 2014) that generates AN firing rates, receives half of the simulated cochlear sections.

The auditory nerve, cochlear nucleus, and inferior colliculus levels are where population responses can be written out. The population AN response sums 13 high-spontaneuos-rate (HSR) (70 spikes/s), 3 medium-spontaneuos-rate (MSR) (10 spikes/s), and 3 low-spontaneuos-rate (LSR) (1 spikes/s) fiber responses r_{AN}. These waveforms (r_{AN}) are then summed to yield the ABR wave-I across each CF and for CFs between 112 Hz and 12 kHz. The ABR wave-I most accurately depict the population's total auditory-nerve activity. In order to produce the ABR wave-III and wave-V, r_{AN} is passed through a same-frequency bushy cell (second order neurons) model, modeling generators for the cochlea nucleus and inferior colliculus, respectively. Simulated W-I, W-III, and W-V waveforms are added and then subjected to a Fourier transform to generate the response component corresponding to the modulation frequency of the EFR stimulus in order to model the EFR.

By altering the cochlea's mechanical gain on a CF-dependent basis, the model can be made hearing-impaired by producing wider cochlear filters that mimic the impact of OHC loss. A model parameter can be set to reflect a BM gain reduction corresponding to a particular hearing sensitivity loss [in dB HL].

To get an idea of how the model works, in particular how hearing loss is simulated, we have generated model outputs (AN,IC and EFR) with simple pure tones (Figure 4 B.). We compared the model settings for the normal hearing human (NH) to a typical mild-to-moderate presbycusis (HI; starting at 1 kHz and sloping to 35 dB HL at 8 kHz). In general, the model exhibits frequency-dependent AN firing rates and increasing amplitudes with rising simulated sound pressure levels. As expected, the models for NH and HI generate similar output for pure tones with F = 500 Hz, since the simulated HL should only affect the processing of frequencies $\leqslant$ 1000 Hz. The sloping HL is reflected in decreasing amplitudes with increasing frequency for 1000 Hz and 4000 Hz. As expected, for higher sound pressure levels, the differences between HI and NH along the simulated model outputs are getting smaller due to the simulated loss of outer hair cells.

2.7 Statistical analysis

This section is not intended to provide a detailed description of the individual statistical methods used in this work. Rather, it aims to present the motivation behind why a particular method was chosen. To address various questions in this thesis, different statistical approaches were employed.

We implemented a novel experimental paradigm using a psychophysically augmented continuous speech paradigm consisting of consecutive trials and short repeats that participants had to detect. We measured several subjects and took multiple measurements from each subject. Mixed models, also known as hierarchical linear models, were used to handle such nested data where observations are not independent (Raudenbush & Bryk, 2002; Gelman & Hill, 2006). The Intraclass Correlation Coefficient (ICC) is used to quantify whether the data were nested (Shrout & Fleiss, 1979).

Unlike in simple regression, model parameters in mixed models are not estimated via ordinary least squares (OLS) but with (restricted) maximum likelihood (ML). The ML method estimates the parameters that maximize the likelihood function, which provides the most likely values of the parameters given the observed data (Pinheiro et al., 2007). Mixed models can model both random and fixed effects simultaneously. Fixed effects are variables that were measured in all the levels of interest (Tukey et al., 1977), while random effects are variables that were measured in only one random level and thus allow for random variation between groups and levels. For example, in our experimental paradigm, trials (repeated measurements from the same subject; level 1) are nested within each individual subject (level 2) and are therefore modeled as random effects.

Random effects can be modelled as a random intercept and/or random slope in the mixed model. Including random slopes in the model may be beneficial if level 2 units vary not only in their average (intercept) but also in their relationship (slope). To test whether including random intercepts or slopes improves the model fit, a model comparison using the Akaike Information Criterion (AIC) is useful. AIC determines which model provides the best fit to the data, balancing goodness of fit and model complexity (Akaike, 1974). Modeling fixed and random effects simultaneously allows for more accurate estimation of model parameters. Additionally, mixed models can handle unbalanced designs and missing data, unlike traditional methods such as ANOVA. A potential source of missing data within this thesis could be due to connection loss during Bluetooth EEG recording or non-responses to embedded repeats in the speech streams.

Overall, mixed models offer a powerful and flexible method for analysing complex data and were therefore predominantly used in this thesis. Mixed models were executed using MATLAB, JAMOVI, and R (The MathWorks, 2021; R Core Team, 2021).

In this thesis, we sought evidence in favour of the null hypothesis. In frequentist hypothesis testing, the p-value provides evidence against the null hypothesis rather than evidence in favour of it (e.g, J. O. Berger & Sellke, 1987). However, Bayesian statistics allow for direct calculation and interpretation of the evidence in support of the null hypothesis (for review, see Rouder, Speckman, Sun, Morey, & Iverson, 2009).

In Bayesian statistics, one determines the distribution of a quantity (posterior distribution) to estimate its true value or test a hypothesis concerning this value. Frequentist statistics assumes that the population parameters are fixed and unknown and that samples are drawn at random from the population, whereas Bayesian statistics does not require random sampling but instead assigns probabilities to the parameters based on prior knowledge or beliefs. This is the main distinction between frequentist and Bayesian statistics.

Bayesian statistics is linked to three different types of probabilities: likelihoods, posterior probabilities, and prior probabilities. We have prior beliefs about the possible values that the statistical model's input parameters might have before observing any data. We express our beliefs as a prior probability distribution, indicating the probability of various parameter values before data are observed. Using a likelihood function given a set of parameter values, we determine the likelihood of observing the collected data. The likelihood function describes how well the model fits the data for a specific set of parameter values.

We revise our prior beliefs about the parameters in light of the data after we observe it. By calculating the posterior probability distribution, which combines our prior beliefs (expressed as the prior probability distribution) and the likelihood function, we accomplish this. In Bayesian statistics, the Bayes factor (Kass & Raftery, 1995) is a measure of the strength of the evidence for one hypothesis over another (see Figure 5 for more details).

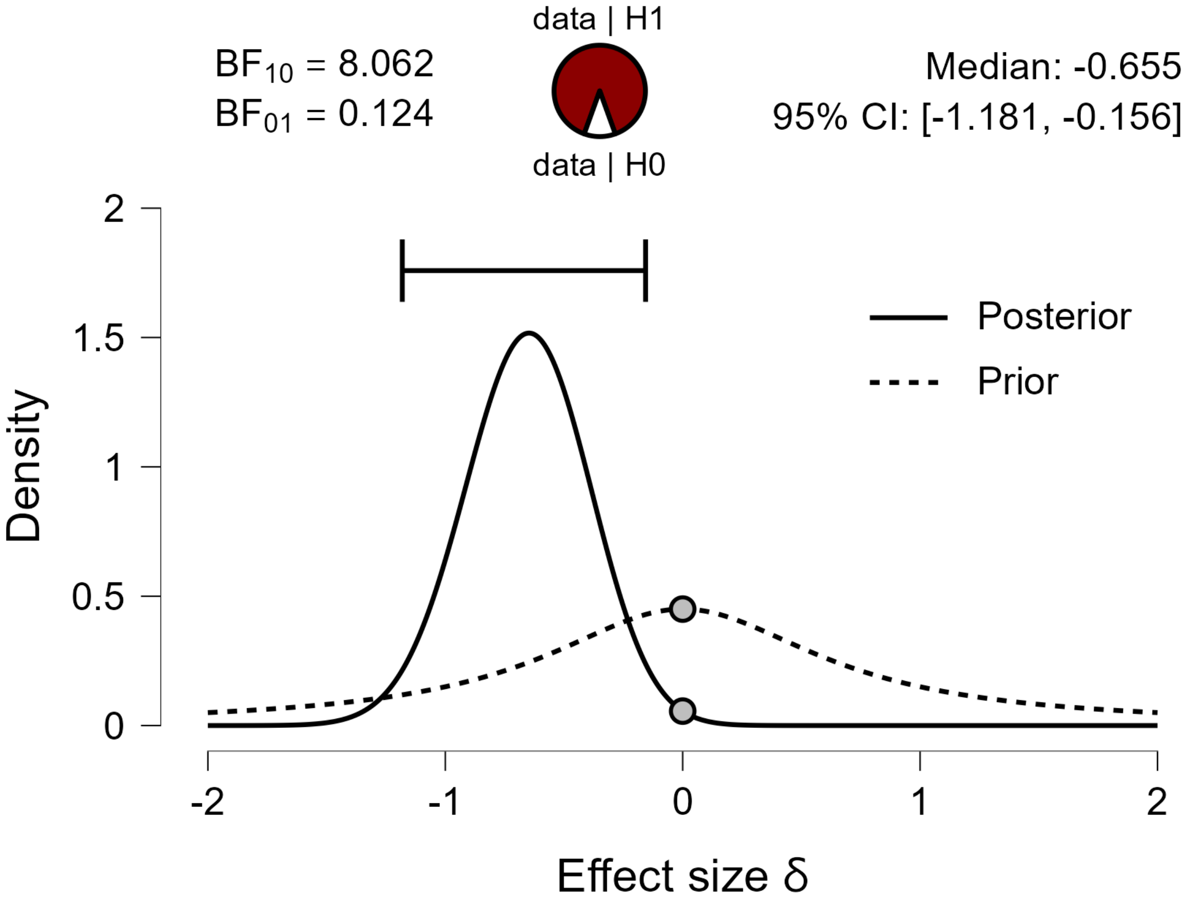


Figure 5: Bayes factor example

Example Bayes factor. Prior (dashed line): We did not expect a difference (similar to null hypothesis), $\delta = 0$. In addition, the prior distribution (Cauchy distribution) also allows for large effects in both directions. The posterior distribution (solid line) shows a narrower distribution after being updated by data, $\delta \approx 0.7$. The Bayes Factor (BF), as illustrated by the so-called pizza plots, is reflecting the likelihood ratio, with which the prior odds (the relative probabilities of the two hypotheses to each other a priori any data; unknown) get updated. BF_{10} indicates the Bayes factor of H1 vs. H0. The red area illustrates that the probability of the data given the alternative hypothesis is 8x larger than the data given the null hypothesis. The key advantage of the BF is that both hypotheses can be compared at the same time.

In the example in Figure 5 the Bayes factor indicates moderate evidence for H1. This qualitative interpretation of the Bayes factor is based on a table proposed by Jeffreys (1998) ranging Bayes factors regarding their evidence. Interestingly, if we assume that we have discovered evidence for H0, the posterior distribution shifts to $\delta \approx 0$ and the white area of the pizza plot becomes prominent. In this thesis, the Bayes factor was used to determine whether the data supported one hypothesis over another. The Bayes factors were computed using JASP and JAMOVI (The jamovi project, 2022, JASP Team, 2023).

To analyse data originating from multichannel EEG measurements, we used cluster permutation tests to investigate differences in time-shifted neural tracking (TRFs) between experimental conditions such as target and neutral distractor. Cluster permutation tests involve clustering together time points of a discrete time series and spatial locations on the scalp, and then testing whether these cluster-level statistics differ significantly between conditions. We used one-sample t-tests to compare the time-series of experimental conditions against zero as a test statistic at the single-subject level. The resulting t-values and a threshold that was set to t-values corresponding to $p < 0.05$ for at least three neighbouring electrodes were used to define clusters at the group level. A permutation distribution of 5000 clusters was used to compare each observed cluster to those clusters. The permutation distribution was generated by randomly assigning the time-resolved experimental data to conditions. P-values were corrected for multiple comparisons using the Monte Carlo method, which takes into account the number of clusters and multiple comparisons across time and space (Maris & Oostenveld, 2007). The cluster p-value represents the proportion of Monte Carlo iterations during which the observed cluster's summed t-statistic is exceeded. The results are interpreted based on these corrected cluster p-values. We performed cluster-based permutation tests using an established two-level statistical analysis implemented in Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011).

To investigate brain-behaviour relationships, we used logistic regression, a variant of generalized mixed model (GMM, J. Fox, 2015; Eid, Gollwitzer, & Schmitt, 2017). We used brain data to predict participants perceptual performance. We modelled behaviour as binary response (hit vs. miss). A simple regression is insufficient to model binary outcome since a binary outcome violates the assumptions of the linear regression model, such as normally distributed errors (Hosmer Jr, Lemeshow, & Sturdivant, 2013). GMMs are characterised by a so-called link function g that is linearly related to the predictors. The logit-function $\text{logit}(p)$ (logarithm of the odds) serves as the link function in logistic regression.

$$g = \text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \sum \boldsymbol{\beta X} \tag{8}$$

where p is the probability of the event occurring, β is the regression coefficient and X the predictor matrix. To predict the outcome, we have to use the inverse link function g^{-1}. However, the regression coefficients β are interpreted in terms of odds ratios. To test the regression coefficients for significance, different statistical tests can be used. Within this thesis, we used the Wald-test that is a statistical hypothesis test that uses the chi-squared test statistic to compare the coefficients of a regression model to a null hypothesis value (Wald, 1943). We used JAMOVI and R to run logistic regression and the Wald-test.

3 Study 1: Auditory neural tracking reflects target enhancement but not distractor suppression in a psychophysically augmented continuous-speech paradigm

3.1 Abstract

Selective attention modulates the neural tracking of speech in auditory cortical regions. It is unclear whether this attentional modulation is dominated by enhanced target tracking, or suppression of distraction. To settle this long-standing debate, we employed an augmented electroencephalography (EEG) speech tracking paradigm with target, distractor, and neutral streams. Concurrent target speech and distractor (i.e., sometimes relevant) speech were juxtaposed, with a third, never task-relevant speech stream serving as a neutral baseline. Listeners had to detect short target repeats and committed more false alarms originating from the distractor than from the neutral stream. Speech tracking revealed target enhancement but no distractor suppression below the neutral baseline. Speech tracking of the target (not distractor or neutral speech) explained single-trial accuracy in repeat detection. In sum, the enhanced neural representation of target speech is specific to processes of attentional gain for behaviourally relevant target speech rather than neural suppression of distraction.

3.2 Introduction

Selective attention refers to the neural filtering processes of prioritizing relevant objects over irrelevant distractions (Desimone et al., 1995). Typically, attentional selection is quantified by the difference in the behavioural or neural response to target versus distractor. However, such a difference can be driven by either target enhancement, distractor suppression, or a combination of the two. Here, we investigated how the mechanism of selective attention is represented in neural (electroencephalographic) activity and we linked the trial-by-trial neural responses to behavioural responses associated with different sub-processes of attention.

In the visual domain, single-cell studies have shown that attention operates when multiple stimuli compete for access to neural representation. Distractors within a receptive field become suppressed, while attended stimuli are enhanced (Desimone et al., 1995). The mechanism of how selective attention is implemented at the level of neural networks is still in debate in attention research (Schneider et al., 2022; van Moorselaar & Slagter, 2020). It has been argued that an often-missing, pre-defined baseline is needed to test whether the target exceeds the baseline (enhancement) and the distractor falls below the baseline (Gundlach, Forschack, & Müller, 2022; Wöstmann et al., 2022). In the visual modality, Seidl et al. (2012) had implemented such a ”neutral” baseline by assigning a given class of stimuli as the never task-relevant, and therefore least distracting, category. They measured brain activity in fMRI (functional magnetic resonance imaging) in response to natural scene photographs that contained objects from a task-relevant (target) category, a task-irrelevant (distractor) category, and a never task-relevant (neutral) category. In addition, distractor suppression was linked to attentional capture. A distractor requires to capture attention initially, followed by suppression (Alexopoulos, Muller, Ric, & Marendaz, 2012; Dalton & Lavie, 2004; Gaspelin & Luck, 2018).

Speech is one of the most salient and behaviourally relevant signals in human environments, but for a long time it was not possible to study the neural processing of time-varying natural stimuli like speech quasi-continuously. Neuroscientists thus studied attention to short, isolated events due to the need for temporally discrete event-related potentials (ERP; Handy, 2005). Recently, research has begun to investigate the electrophysiology of attention to continuous speech (Ding & Simon, 2012; Lalor & Foxe, 2010; Wöstmann, Fiedler, & Obleser, 2017). Electrophysiological responses in cortical regions phase-lock to the temporal envelope of the speech signal (Luo & Poeppel, 2007). This linear relationship is well-captured by the so-called temporal response function (TRF), which can be interpreted as a cortical impulse response, in close analogy to the conventional ERP (Crosse et al., 2016; Fiedler et al., 2019). The TRF can indicate a stereotypical, phase-locked brain response to various acoustic features. The most often used feature is the low-frequency temporal envelope, also referred to as neural speech tracking (Obleser & Kayser, 2019). This neural speech tracking shows a robust and often-reproduced differentiation of attended versus ignored speech (Ding & Simon, 2012; Fiedler et al., 2019; Horton et al., 2013; Kerlin et al., 2010; Mesgarani & Chang, 2012). Thus, neural tracking is a feasible approach to quantify the neural processing of several speech streams at the same time to reveal the effect of attention (Ding & Simon, 2012; Puvvada & Simon, 2017; Golumbic et al., 2013). In addition, Fiedler et al. (2019) showed that late TRF components are associated with cortical tracking of ignored speech and are differently modulated for varying signal-to-noise ratios. These findings indicate that different components of the TRF are associated with different attentional processes. In sum, a hitherto underutilised advantage of this approach is its ability to delineate two potential sub-processes of attention: target enhancement vs distractor suppression (Wöstmann et al., 2022).

What characterises a distractor stream in such an experimental setup? First, the implementation of the distractor stream was based on the phenomenon of ”negative priming,” which describes the finding that a distractor from the previous trial is harder to select on the next trial (Kristjánsson & Driver, 2008; Shiffrin & Schneider, 1977; Tipper, 1985) . It is assumed that a stimulus and the response it elicits become integrated into so-called ”event files” in memory (Frings et al., 2015). Therefore, a specific stimulus automatically retrieves the response that was previously linked with this stimulus (Hommel, 1998). In this sense, the whole distractor stream in a given trial is distracting, since the same event that was previously task-relevant triggers a response despite currently being task-irrelevant, and must be inhibited.

Second, it was shown that spatial statistical regularities influence selective attention on a longer time scale. A location that contained a distractor with a higher probability is suppressed relative to other locations. In this context, participants would learn about the location of the distractor stream and suppress it over time (Wang & Theeuwes, 2018b).

In the auditory modality, (Hambrook & Tata, 2019) investigated the mechanisms of distraction by increasing the number of distractor streams in the auditory scene. Their results suggest that distraction is not an active process but rather simply a loss of phase tracking of the target envelope. However, the attentional sub-processes target enhancement and distractor suppression have been suggested but have rarely been probed explicitly (Fiedler et al., 2019; Petersen et al., 2017; Vanthornhout, Decruy, & Francart, 2019). We followed Seidl et al. (2012) logic and implemented three auditory speech streams: a target (task-relevant) stream, a distractor stream (previously task-relevant), and a neutral stream that is never task-relevant. Larger target-vs-neutral tracking would indicate enhancement, while smaller distractor-vs-neutral tracking would indicate suppression. In the context of the auditory scence, the neutral stream can be conceived as a weaker distractor not as non-distractor. We operationalized the neutral stream as the never task-relevant stimulus. However, the neutral stream is not neutral in the strongest sense: Like the distractor stream, it was associated with the attentional background since it had to be ignored by the listener (Puvvada & Simon, 2017). In other words, the neutral stream was more similar to the distractor stream compared to the target stream. Critically, it is conceivable that suppression is preceded by initial attentional capture of the distractor, indicated by larger distractor-vs-neutral tracking for early neural responses (see Fig. 6B).

However, a severe disadvantage of continuous speech paradigms thus far has been their typical lack of rich behavioural data (Hamilton & Huth, 2020). Typically, comprehension questions are asked intermittently or afterwards regarding the content of the audio stream, which are insufficient to assess the task-relevance of neural responses, especially during a complex continuous speech paradigm.

In the present study, we use electroencephalography (EEG) to investigate neural responses in human participants. We asked to what extent selective attention to speech is implemented in the human brain through target enhancement versus distractor suppression, and whether en-hanced tracking of target speech or suppressed tracking of distraction would explain behav-ioural trial-by-trial indices of selective attention.

To this end, we designed a new experimental paradigm with two key advances over previous neural speech-tracking experiments (Fig. 6A). First, a speech stimulus that was never relevant served as a neurally and behaviourally ‘neutral’ baseline, against which the processing of concurrent target speech (relevant on a present trial) and distractor speech (relevant on other trials) can be contrasted (Seidl et al., 2012; Wöstmann et al., 2022). Second, listeners had the task of continuously monitoring and detecting short repeats in the target stream (Marinato & Baldauf, 2019) and ignoring short repeats in the distractor and neutral streams. This enabled us to contrast whether neural responses to target, neutral, or distractor speech would independently explain trial-by-trial variation in attentional performance.

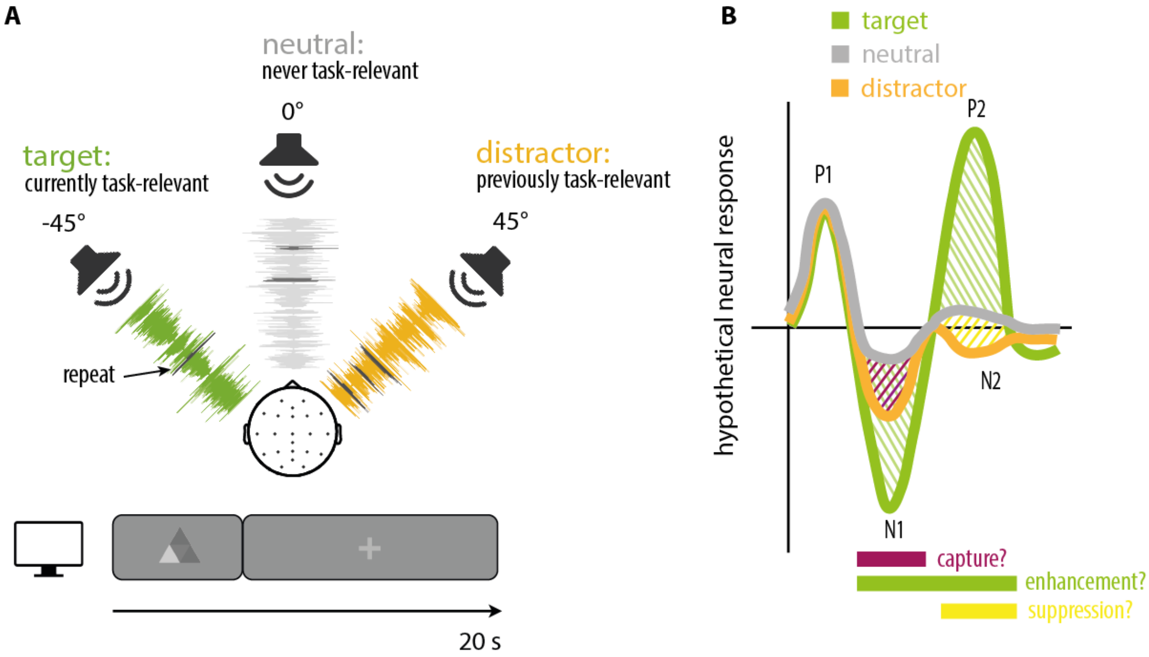


Figure 6: Experimental design and hypothetical results

Experimental design and hypothetical results. A. Simultaneously, we presented three different audio streams at different locations (-45°, 0°, 45°). Participants were instructed to attend to the cued audio stream for the duration of a trial (currently task-relevant target). In the next trial, another stream was cued, which became the target stream. The stream that was previously task-relevant became the distractor stream. During the entire experiment, the cue alternated between these two streams. The task-irrelevant (never cued) stream was defined as the neutral stream. In all three streams, we included short repeats. Participants had to detect repeats in the target stream and ignore repeats in the neutral and distractor streams. Further, participants were instructed to process the content of the target audio stream. **B.** Hypothetical neural outcomes. While target enhancement (stronger target vs. neutral tracking; green) is expected for early and late TRF components, earlier components are expected to show neural capture by the distractor, that is, distraction (stronger distractor vs. neutral tracking; red), and later components are expected to show suppression (reduced distractor vs. neutral tracking; yellow).

3.3 Methods

3.3.1 Participants

The current study included 19 young adults (12 females and 7 males) ranging in age from 18 to 27 years (21.9). All participants had German as their mother tongue and reported normal hearing and no histories of neurological disorders. To verify normal hearing, we measured pure-tone audiometry within a range of 125 to 8,000 Hz. All participants showed auditory thresholds below 20 dB HL for the tested frequencies. They gave written informed consent and received compensation of 10 euros per hour. The study was approved by the local ethics committee of the University of Lübeck.

3.3.2 Stimulus materials and spatial cue

We presented three different narrated book texts as audio, spoken by different male, untrained talkers (”Michael Kohlhaas” by Heinrich von Kleist, ”Pole Poppenspäler” by Theodor Storm, and ”Das Wrack” by Friedrich Gerstäcker). We chose audio streams that were fictional instead of fact-based, to minimise the impact of variations in prior knowledge on a topic and a resulting possible bias to one of the audio streams. At an SPL (mixture) of about 65 dB(A), which corresponds to normal conversation levels, all three audio streams overlapped in time.

The following processing steps of the stimuli were done using custom written code in MATLAB (Version 2018a Mathworks Inc., Natick, MA, United States). The sound files were sampled at 44.1 kHz with a 16-bit resolution. The sound level was matched to the same long-term root-mean-square (rms) dB full scale (dBFS) between the three audio streams. Silent periods were truncated to maximally last 500 ms (O'Sullivan et al., 2015).

We embedded short repeats in the audio streams by pseudo-randomly selecting a 400-ms segment from the original stream and repeating it directly thereafter (Marinato & Baldauf, 2019). The first repeat was presented at least two seconds after stimulus onset. A linear ramping and cross-fading technique was used to incorporate each repeat into the sound stream. The linear ramping was done by using a window of 220 samples (5 ms) at the end of the part to be repeated (the down ramp) and the first 220 samples (5 ms) of the repeat itself (the up ramp). The cross-fading was done by adding the down and up ramps together.

The onset time of each repeat was drawn randomly to avoid predictability of the repeat. To avoid that repeats occurring in the different streams overlap in time, the distance between two repeat onsets was at least 2 seconds.

We further used a rms (root mean square) criterion (the rms of the repeat had to be at least the same as the rms of the stream from which the repeat was drawn) to avoid undetectable repeats of low sound intensity.

The cue was presented at the center of the screen (resolution: 1920x1080, Portable HDMI Screen, Wimaxit) in front of the participant (distance: 1 m). The cue (Fig. 6A) consisted of three sub-triangles that had a size of 1.3° and pointed to the three sound sources (front, left, and right). The background of the screen (RGB: 127, 127, 127), the cued sub-triangle (RGB: 204, 204, 204), and the not cued triangles (RGB: 115, 115, 115) were kept in different shades of gray to keep the contrast low. The bright triangle indicates the to-be-attended position. Since the cue and the fixation cross were presented at the same time as the auditory stimuli, we ensured that the possible interference between visual and auditory neural responses was as small as possible. To this end, the change between the fixation cross and cue was made smooth by linearly fading in and fading out (50 ms each) the cue.

3.3.3 Experimental Setup

The experiment took place in a laboratory space with eight loudspeakers (Genelec: Speaker 8020D, Denmark) arranged in a circle with a radius of one meter. The loudspeakers were spaced at 45 degrees. A chair was placed in the middle of the radial speaker array, face-aligned to the loudspeaker at position 0°. The three audio streams were presented over the three frontmost loudspeakers (-45°, 0°, and 45° in the azimuth plane, elevation was not adjusted for participants' height, ground-to-loudspeaker distance: 1,20 m, the five remaining speakers were not used in the present experiment). In advance, participants were briefed about the experiment. Importantly, they were not briefed about the condition-to-location assignment of the streams. Each participant was asked to keep their eyes open, focus on the center of the screen, and sit as relaxed as possible. To avoid head motion, a chin rest was used. The height of the chin rest was adjusted for each participant.

Each participant had to switch their attentional focus between the same two streams and locations. The stream at the cued location was defined as target, the stream cued in the previous trial was defined as distractor. Importantly, this left each participant with only one, never task-relevant stream and location, here defined as neutral. Between participants, we implemented three condition-to-location assignments to avoid any confound with the position of the neutral stream (neutral front (0°), neutral left (-45°), and neutral right (45°). We aggregated across the three condition-to-location assignments to obtain our measures of interest, i.e., neutral tracking of target, neutral, and distractor. As the position of the neutral stream, the different audio streams were almost balanced between the 19 participants (neutral front: n=7; neutral right n= 6; neutral left n= 6).

Participants had to detect short repeats in the target stream. Each trial contained 6 repeats, which were randomly partitioned into the three streams (for procedure details, see section: Stimulus materials and spatial cue) Before data collection, participants were familiarized with the experiment. During instruction, it was emphasized to respond as fast and accurately as possible to a repeat in the target stream, but also to listen to the content of the target stream. To familiarize participants with the repeats, we presented them with a single sentence with one repeat included. They had to give oral feedback if they were able to detect the repeat. Further, we presented them with 6 training trials that corresponded to the main experiment but used different audio streams. The main experiment consisted of 180 trials divided into 4 blocks, resulting in a total duration of 60 min. After each block, participants were able to take a rest. The total number of repeats was 360 per stream across the experiment.

We asked participants 15 multiple choice questions (with four possible answers, each) about the content of each audio stream at the end of the experiment. To avoid participants attending the to-be-ignored audio stream, we did not ask the questions after every block. The order of the questions and the possible answers were randomized between participants.

3.3.4 Behavioural data analysis

We evaluated participants' behavioural performance in two ways. We analysed the proportion of detected repeats and, as a control, the proportion of correctly answered content questions.

We analyzed the detection of repeats in terms of signal detection theory. Button presses to repeats in a time window (150-1500 ms) after repeat onset were considered in this analysis. A button press following a repeat in the target stream was assigned as a hit. Button presses following repeats in the distractor stream and in the neutral stream were assigned as separate types of false alarms. To differentiate between false alarms to repeats in the neutral versus distractor stream, we calculated sensitivity (d') between hit rate and false alarms to distractor repeats [$d'_{target\,vs.\,distractor} = z(hit\,rate)–z(false\,alarm\,rate\,distractor)$] and hit rate and false alarms to neutral repeats [$d'_{target\,vs.\,neutral} = z(hit\,rate)–z(false\,alarm\,rate\,neutral)$]. For this signal-detection analysis of repeats, we excluded one participant who did not respond to any repeats in the distractor stream.

A challenge in creating multiple-choice comprehension questions is to provide multiple (here: four) response options that cannot be solved based on prior knowledge or the possibility of excluding some of the response options. Hence, participants' actual guess rate might be considerably higher than the theoretical chance level of 25%. Thus, in a pilot experiment, we presented the multiple-choice comprehension questions to N=9 different participants who had not listened to the audio streams at all. As a result, a new "empirical" chance level of 40% (3.9% S.E.M.) was established. In the following, we tested the proportion of correctly answered questions in the main experiment against this empirical chance level.

3.3.5 Data acquisition and pre-processing

EEG was recorded using a 24-electrode EEG-cap (Easycap, Herrsching, Germany; Ag–AgCl electrodes placed according to the 10-20 International System) connected to a SMARTING amp (mBrainTrain, Belgrade, Serbia). This is a mobile EEG system, that transfers the signal via Bluetooth to a recording computer (e.g., Waschke, Wöstmann, & Obleser, 2017; Wöstmann, Waschke, & Obleser, 2019). EEG activity was recorded with the software Smarting Streamer (mBrainTrain, version: 3.4.2) at a sampling rate of 500 Hz. During recording, electrode FCz served as an online reference, and impedances were kept below 20 kΩ. No data loss was reported during the sessions.

Offline, EEG preprocessing was done using MATLAB (Version 2018a, Mathworks Inc., Natick, MA, United States), built-in functions, custom-written code, and the Fieldtrip-toolbox (Oostenveld et al., 2011). EEG data were re-referenced to the average of M1 and M2 (left and right mastoids) electrodes and high- and low-pass filtered between 1 and 100 Hz (two-pass Hamming window, FIR). An independent component analysis (ICA) was computed on each participant's EEG data. M1 and M2 were removed before ICA. ICA components related to eye blinks, eye movement, muscle noise, channel noise, and line noise were identified by visual inspection and removed. On average, 8.37 of 22 (SD = 3.13) components were rejected. Components that were not associated with artifacts were projected back into the data. Clean EEG data were further processed. Hence, EEG data were low-pass filtered again at 10 Hz (two-pass Hamming window, FIR). Afterwards, EEG data were resampled to 125 Hz and segmented into epochs corresponding to the 20-s trial length.

3.3.6 Extraction of the speech envelope

The temporal fluctuations of speech were quantified by computing the onset envelope of each audio stream (Fiedler et al., 2017). First, we computed an auditory spectrogram (128 sub-band envelopes logarithmically spaced between 90-4000 Hz) using the NSL toolbox (Chi et al., 2005). Second, the auditory spectrogram was summed up across frequencies, resulting in a broadband temporal envelope. Third, the onset envelope was obtained by computing the first derivative of this envelope and zeroing negative values to obtain the half-wave rectified first derivative. Finally, the onset envelope was downsampled to match the target sampling rate of the EEG analysis (125 Hz). Compared to the envelope, using the onset envelope shifts the envelope in time. Importantly, the TRF obtained by using the onset envelope as a regressor has the most similarity to a classical ERP (Fiedler et al., 2017).

3.3.7 Temporal response functions (TRFs)

The deconvolution kernel or impulse response, which describes the linear mapping between an ongoing stimulus and an ongoing neural response, is called the temporal response function (TRF). We used a multiple linear regression approach to compute the TRF (Crosse et al., 2016). More precisely, we trained a forward model using the onset envelopes (e.g., Fiedler et al., 2019) of the target, distractor, and neutral speech to predict the recorded EEG response. In this framework, we analysed time lags between –100 and +500 ms between envelope changes and brain responses.

To account for the EEG variance attributable to the detection and processing of the behaviourally relevant repeats and corresponding evoked brain responses, we also included all onsets of the repeats in the three streams and the button press in the model as nuisance regressors, represented by stick functions. The onsets of the repeats are independent of the speech envelope regressors by design, since these were almost randomly (within the constraints of SNR threshold) added into the speech streams.

To prevent ill-posed problems and overfitting, we used ridge regression to estimate the TRF (Crosse et al., 2016). Lambda (λ) is the ridge parameter for regularization. We estimated the optimal ridge parameter that optimized the mapping between stimulus and response by leave-one-out cross-validation for each participant. First, the stimuli are segmented in M-trials and different ridge values ($\lambda = 2^0, 2^1, \dots 2^{20}$) are predefined. In this approach, a separate model for each λ is calculated. Second, the trials are mixed, and each time one is left out. This trial is used as a test set, while the M-1 trials are used as a training set. Then, the models are averaged over the trials and convolved with the data from the matching test set to predict the neural response. This is done for every predefined λ. Calculating the MSE between the predicted estimate and the original data provides a validation metric that enables selecting the λ with the lowest MSE. We used the ridge value with the lowest MSE (specific for each subject) for the TRF model that jointly contained the target, distractor, and neutral onset envelopes as regressors.

TRFs were estimated based on the trials in the experiment. Participants had to switch their attention trial-wise between two of the streams. Hence, the trials enable the assignment of target, distractor, and neutral onset envelopes. The time window in which the stimulus and response are cut to estimate the TRF is referred to as a "trial." To avoid any conflicts with the cue, the first second of each trial was cut off in the EEG signal and the envelope onsets. One model was trained on 180 trials, incorporating multiple predictor variables: the onset envelope for target, distractor, and neutral streams; and the stick functions for the repeats and button presses. Resulting in a single TRF for each predictor variable that predicts a separable response component. Similar to the TRF approach, we estimated TRFs for the embedded repeats, but we modelled repeats as a stick function based on the repeat onset. Importantly, TRFs for the three streams, TRFs for repeats in the three streams, and button presses were estimated in the same model with the same regularization.

3.3.8 Neural tracking

Neural tracking quantifies how strongly a single stream is represented in the EEG signal. TRFs were used to predict the EEG response. The neural tracking (r) was calculated by correlating the predicted and measured EEG responses using Pearson correlation. We predicted the EEG signal on single trials using the leave-one-out cross-validation approach (see above). The r-values that resulted were averaged across trials and participants. We obtained the neural tracking accuracy over TRF time lags by using a sliding-time window of time lags (size: 48 ms, 6 samples) with an overlap of 24 ms (3 samples) for the prediction (Fiedler et al., 2019; Hausfeld, Riecke, Valente, & Formisano, 2018; Kraus, Tune, Ruhe, Obleser, & Wöstmann, 2021; O'Sullivan et al., 2015). For every window position, the neural tracking was calculated, resulting in a time-resolved neural tracking. We used the term "stream tracking" which refers to the neural tracking of the envelope onsets, and "repeat tracking," which refers to the neural tracking of the repeat onsets. To obtain the repeat tracking, we used the same pipeline as for the speech tracking procedure (see above), with the exception that we estimated neural tracking based on the onsets of repeats (instead of the speech onset envelope), which we modelled as stick functions.

3.3.9 Statistical analysis

A study (Fiedler et al., 2019) investigated the attentional effects of neural tracking in a comparable continuous speech paradigm by recording the EEG of N = 18 participants. It is reasonable to expect that similar effect sizes will be observed in a replication of auditory attention effects with the same sample size. The present study is supposed to detect neural tracking effects with at least medium to large effect sizes (Cohen's d ≥ 0.7) and a power of 80 % (two-sided, within-subject tests, Alpha = 0.05) for N = 18 subjects.

We also used different statistical procedures to answer different questions. To answer the main research question (outlined in Fig. 6B), we used generalized mixed models (jamovi 1.6, R 4.0). This approach enables us to include and jointly model factors that potentially influence behaviour and the neural response. These included at least the factor condition-to-location assignment (neutral front, left, or right) and the subject as a random intercept to account for between-participant variability.

To determine statistically significant differences in behavioural sensitivity (outcome measure), we included target versus distractor and target versus neutral as categorical predictors in the model. To determine statistically significant differences in neural tracking (outcome measure), we included the target, neutral, and distractor streams as categorical predictors in the model. In both models, we included the factor condition-to-location assignment as a covariate and the random intercept (subject ID) into the model. Bayesian t-tests were calculated to obtain Bayes factors to quantify evidence for the null hypothesis (JASP Team, 2022).

For quantifying the brain-behaviour relations, we used a generalized linear mixed-effects model (repeat detected or not; binomial distribution, with logit link function), since we predicted a binary outcome. The predicted outcome variable was the binary response to the detection of a single repeat in the target stream (Hit = 1; Miss = 0). We included the encoding accuracies for the target, neutral, and distractor streams as continuous, z-scored, fixed-effects predictors in our model. We assigned repeat tracking (trial-based) to each repeat within a trial. To again control for potential confounding between stream tracking and repeats, we also included repeat tracking similar to stream tracking in our model. Beside the factors condition-to-location assignment and subject as random intercepts, we also included the number of repeats during the total experiment and the number of repeats within a trial, as well as the trial number as a random intercept, into the model.

3.3.10 Statistical analysis on time series

We were looking for time points in time-resolved neural tracking that might differ between subjects (target enhancement: neutral vs. target, and active suppression: neutral vs. distractor). To answer this question, we used an established two-level statistical analysis, more specifically a cluster permutation test implemented in Fieldtrip (Oostenveld et al., 2011). Data from 22 channels was used in this analysis. As a test statistic at the single-subject level, we used one sample t-tests to test the time-resolved neural tracking to the target, neutral, and distractor as well as the neutral-target, neutral-distractor, and target-distractor difference against zero. At the group level, clusters were defined by the resulting t-values and a threshold that was set to t-values that corresponded to $p < 0.05$ for at least three neighboring electrodes. Each observed cluster is compared to 5000 clusters with a permutation distribution. The permutation distribution was generated by randomly assigning the time-resolved neural tracking data to conditions. The Monte Carlo method was used to correct for multiple comparisons. The relative number of Monte Carlo iterations in which the summed t-statistic of the observed cluster is exceeded is indicated by the cluster p-value (Maris & Oostenveld, 2007).

3.4 Results

We recorded the electroencephalogram (EEG) from 19 young, normal-hearing participants (7 male and 12 female, mean age 21.9 years, range 18–27 years). They were presented with three continuously narrated audio streams simultaneously (Fig. 6A). On a trial-by-trial basis, they had to switch their attention between the same two audio streams. The to be attended audio stream was defined as the target stream, the audio stream attended in the trial before as the distractor stream, and the never task-relevant audio stream as the neutral stream. Participants had to detect any repetitions in the target stream as fast and accurately as possible and ignore the neutral and distractor streams.

Here, we analysed behavioural data in terms of signal detection theory. We tested whether selective attention is driven by an enhancement of the target, a suppression of the distractor, or a combination of the two by investigating the differential neural tracking of target versus neutral speech and distractor versus neutral speech by slow (1-8 Hz) cortical responses.

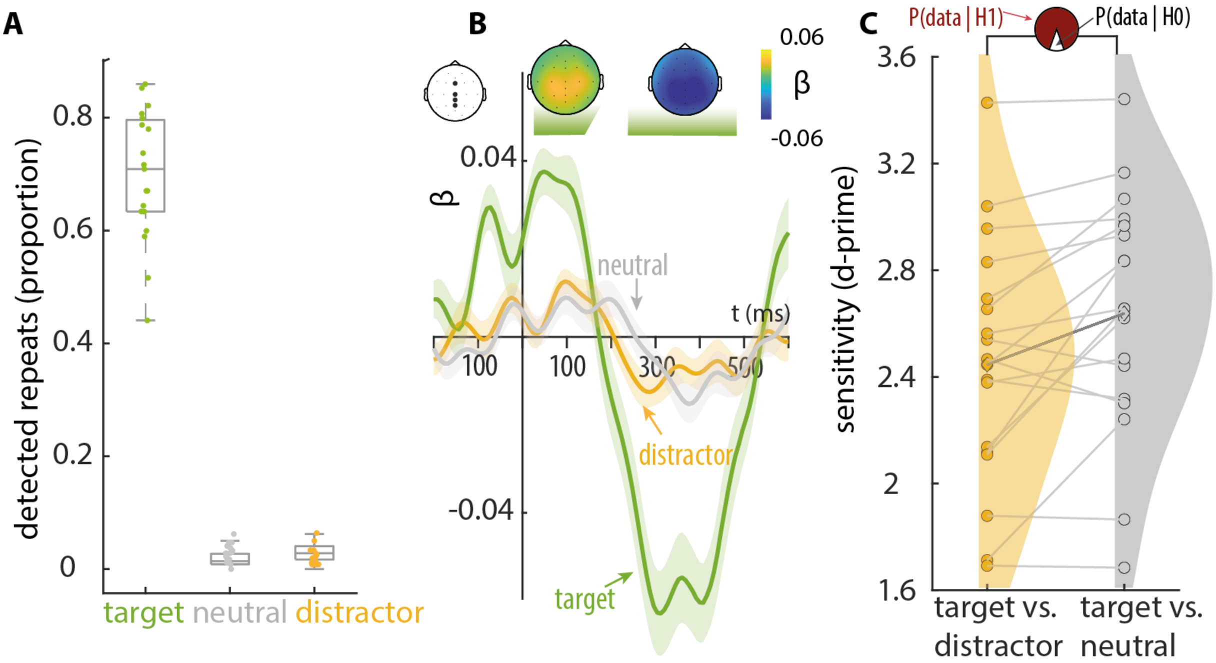


Figure 7: Behavioural results and ERPs to repeats

Behavioural results and TRFs to repeats. A. Box plots depict the proportion of detected repeats for the target (green), neutral (gray), and distractor stream (orange). Scatter dots depict individual subject data. **B.** TRF to repeats in the target (green), neutral (gray) and distractor stream (orange). TRFs ß-weights are averaged across subjects (N=19) and channels of interest (solid line). Shaded areas show the standard error for each time lag across subjects. Topographic maps depict ß-weights for an early time window (0-100 ms) and for a later time window (300-400 ms) for the attended stream.**C.** The spaghetti plot shows the sensitivity index (d-prime) for target versus distractor streams and target versus neutral streams. Dots depict individual data, with connection lines indicating data from the same subject. Shaded areas illustrate the distribution of the data. Bayes factor visualisation: probability pie charts show the ratio of the likelihood of H1(red) and H0 (white) for pairwise comparisons.

3.4.1 Larger repeat evoked responses in the target stream

Overall, participants were well able to detect repetitions in the target stream (mean accuracy: 69.8% ± SEM 2.7%; response time: 735 ms ± SEM 14.1 ms), but performance was clearly not ceiling up with up to 86% (the highest score of single individual) correct responses (Fig. 7A). In comparison, the false alarm rates for the neutral (false alarm rate: 2.1% ± SEM 3.4%; response time: 789 ms ± SEM 44.7 ms) and distractor streams (false alarm rate: 2.9% ± SEM 3.4%; response time: 801 ms ± SEM 37.7 ms) were low. Jointly, the number of hits and false alarms indicated that participants were attending to the cued target audio streams. No significant differences in response times were observed (t = 2.20; df = 30; $p > 0.05$, for all comparisons).

We also estimated regression-based TRFss phase-locked to repeat onset (Fig.7B). TRFs to repeats in the target stream yield an auditory ERP-typical, biphasic response with an early positive deflection (0-170 ms) and a later negative deflection (170-550 ms). Topographies show ß-weights with the highest magnitude for central channels. In contrast, the TRFs for the neutral and distractor streams did not show clear TRFs. Regression based ERPs indicated a different brain response to target versus neutral repeats, but no different brain response to repeats in the neutral and distractor streams, which is in line with the observed behaviour in Fig. 7A. For further neural analysis, we treated the magnitude of these TRFs in all three streams as potential confounds and controlled for them statistically (for details see Methods: Temporal response functions (TRFs). The fact that the participant's performance was off ceiling for detected repeats in the target stream and had a low false alarm rate in combination with no TRFs to the distractor and neutral streams indicate that the repeats did not pop out of the streams automatically. However, we label repeats "detected" in the distractor and neutral streams (false alarms) only if they are followed by a response. We cannot exclude the possibility that some repeats are detected but not followed by a response (response inhibition), even though TRF for false alarms indicates no pop-out.

3.4.2 Larger interference by distracting versus neutral speech

To better understand the contrast in behaviour between the neutral and distractor streams, we analysed the behavioural data in terms of signal detection theory. Based on the hit rate and false alarm rates, two different d' could be calculated (Fig. 7C). We calculated $d'_{target\, vs.\, distractor}$ to index the perceptual separation of target versus distractor stream, and $d'_{target\, vs.\, neutral}$ to index the perceptual separation of target versus neutral stream. Participants achieved a mean $d'_{target\, vs.\, distractor}$ of 2.46 ± 0.1 (M±SEM) and a somewhat higher mean $d'_{target\, vs.\, neutral}$ of 2.66 (±0.1).

A mixed model (supported by a Bayesian paired-samples t-test) with the regressor attention (target-distractor vs. target-neutral) confirmed this difference to be statistically significant (t = 3.01; df = 15; p = 0.009; BF10 = 8.1, supporting H1 over H0), indicating larger interference by the distractor than the neutral speech stream.

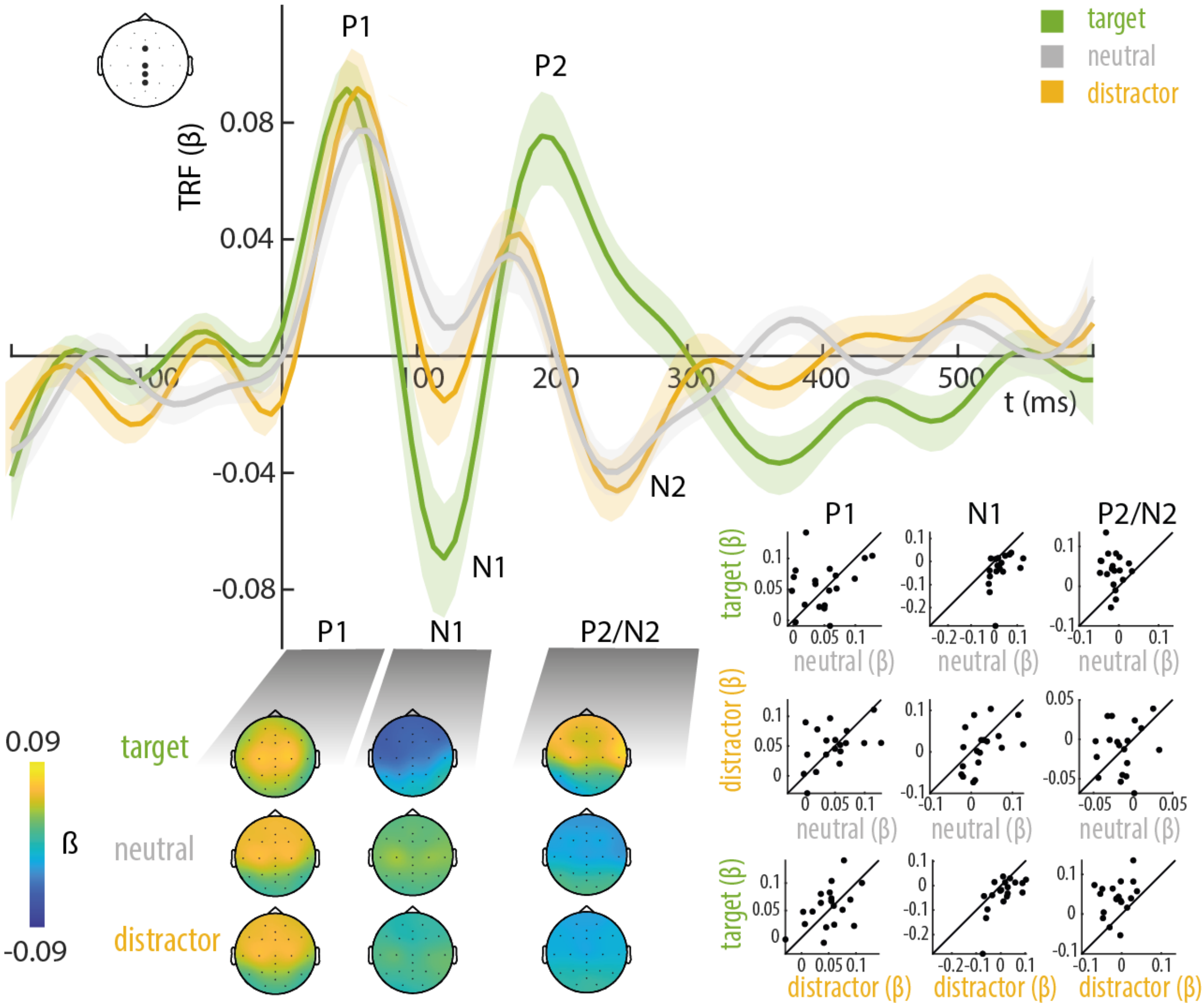


Figure 8: TRF Topo

Temporal response functions (TRFs) of the target, neutral and distractor streams. TRF ß-weights are averaged across subjects (N=19) and channels of interest: Fz, Cz, CPz and Pz (solid lines). Shaded areas show the standard error for each time lag across subjects. Topographic maps depict ß-weights for time windows of the P1, N1 and P2/N2 components for the three streams. 45°-plots show the single subject (N=19) ß-weights separately for neutral versus target, neutral versus distractor, and distractor versus target for the P1, N1 and P2/N2 components.

3.4.3 Morphology of neural responses to target, neutral, and distractor speech

We analysed the neural tracking response to the target, neutral, and distractor streams by investigating the temporal, time-lagged relationship between the stimulus representation of each stream and the brain signal. This relationship is captured by an impulse response, the so-called temporal response function (TRF; see methods). Each component of the TRF is interpreted as a neural operation along the auditory pathway, analogous to the event-related potential (Davis & Johnsrude, 2003; Di Liberto, O'sullivan, & Lalor, 2015). Here, we describe differences between the TRF for the target, neutral, and distractor streams, followed by a statistical analysis of the neural tracking response.

As expected, the morphology of the TRF for the target stream showed the succession of P1-N1-P2 response components, and the TRFs for the neutral and distractor streams showed the succession of P1-N2 response components (Fig. 8).

The early positive deflection P1 (0-80 ms) appeared in the TRF for the target, neutral, and distractor streams without any difference, indicating no attentional modulation. Topographies (located in fronto-central regions), latencies, and polarity of the P1 component were in line with previously observed TRFs and auditory evoked potentials (AEPs) in the literature.

The later negative component N1 (80-150ms) was prominent for the TRF of the target stream. The magnitude of N1 was increased (i.e., more negative) compared with the neutral and distractor streams.

The late positive deflection P2 (170 -300 ms) was only present for the TRF of the target stream. In contrast, we found a negative deflection N2 in the TRF for the distractor and neu-tral stream in about the same time interval. This anti-polar relationship was also reported in previous studies (Ding & Simon, 2012; Fiedler et al., 2019). However, there was no considerable difference in N2 for the TRF of the neutral stream versus the TRF of the distractor stream.

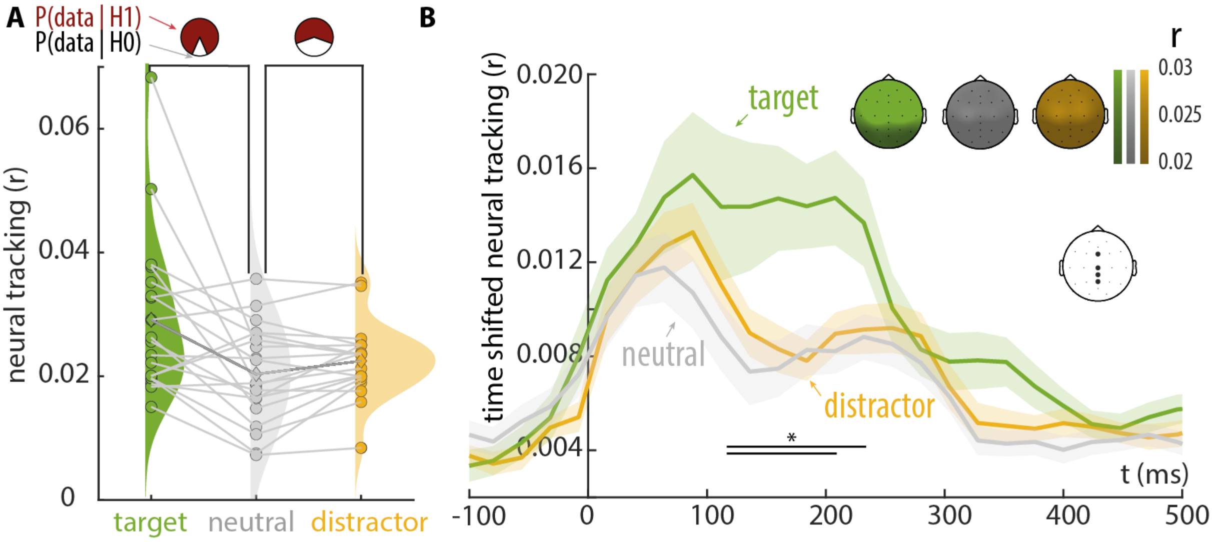


Figure 9: Neural tracking

Neural tracking reveals target enhancement but no distractor suppression. A. Neural tracking was computed based on the extracted TRFs and the envelopes of the attended (green), neutral (gray) and distractor streams (orange). Spaghetti plot shows single-subject data averaged across channels of interest. Connection lines between dots indicate the same subject. Bayes factor visualisation: pie charts show the probability of data given H1 (red) and H0 (white) for pairwise comparisons. Shaded areas depict distributions of the data. **B.** Unfolding neural tracking across time lags (-100-500 ms). Solid lines show the averaged neural tracking (encoding accuracy; r) across subjects (N=19) and channels of interest (topographic map). Shaded areas show the standard error for each time lag across subjects. Cluster permutation test revealed two significant clusters between target and neutral (136-232 ms) and between target and distractor (136-208 ms). Black bars indicate significant clusters. No significant clusters between distractor versus neutral were found. Topographic maps depict average neural tracking (r) for the three streams (0-500 ms).

3.4.4 Neural tracking reflects target enhancement, not distractor suppression

Neural tracking reflects the strength of the representation of a speech stream in the EEG (see methods for details). For neural tracking, we asked whether selective attention is driven by an enhancement of the target, a suppression of the distractor, or a combination of the two. The most important finding of this study resulted from the differential neural tracking of the target and neutral streams (target enhancement; Fig. 9B).

Analysis of the neural tracking (0-500 ms) revealed a difference between the target and neutral stream indicated by a linear mixed model on the mean neural tracking (0-500 ms) and Bayesian t-test for target stream versus neutral stream (t = 3.67; df = 32; p ¡ 0.001; BF10 = 6.5, supporting H1 over H0) and between the target and distractor stream (t = 2.78; df = 32; p¡ 0.05; BF10 = 2, weakly supporting H1 over H0). There was no significant difference in neural tracking of the distractor versus neutral stream and also the Bayes factor is not evidential (t = 0.88; df = 32; p = 0.383; BF10 = 1.6). If at all, the Bayes factor indicates the unexpected finding that the distractor stream was tracked slightly better than the neutral stream (see Fig 9). Topographies revealed the strongest neural tracking for central and frontal channels.

Lastly, we analysed the temporally resolved dynamics of target enhancement and distractor suppression (Fig. 9B). Unfolding neural tracking across time lags revealed differential tracking of the target and neutral streams. Target enhancement of encoding target versus neutral speech was signified by one cluster (136-232 ms; cluster p = 0.0044) We observed no significant clusters separating the neural response to neutral versus distractor stream.

Altogether, these findings indicate that neural tracking in a continuous speech tracking para-digm reflects a neural mechanism of target enhancement at the auditory cortical level, but no active distractor suppression.

3.4.5 Neural tracking of the target stream is associated with perceptual performance

To test the relationship between neural tracking and repeat detection performance, we modelled binary response behaviour (hit vs. miss) as a linear function of neural tracking for the speech streams in the target, neutral and distractor streams using a generalized linear mixed model (GLMM; see methods for details). Further, we also controlled for the different numbers of repeats in the target stream by adding the trial number as a continuous predictor into the model. We also included the subject ID, the number of repeats (total experiment), and the condition-to-location assignment (neutral front, left, right) as random intercepts into the GLMM.

Neural tracking of continuous speech of the target stream displayed a positive linear relationship with participant's performance (β±SEM = 0.077±0.029; z = 2.618; p = 0.009). The higher the tracking accuracy of the target stream during a 20-s trial, the more likely participants detected repeats in that stream during that trial. We observed no such linear relationship in the neutral (β±SEM = 0.017±0.029; z = 0.589; p = 0.556) or distractor streams (β±SEM = –0.023±0.029; z = –0.806; p = 0.420; Fig. 10A, left panel).

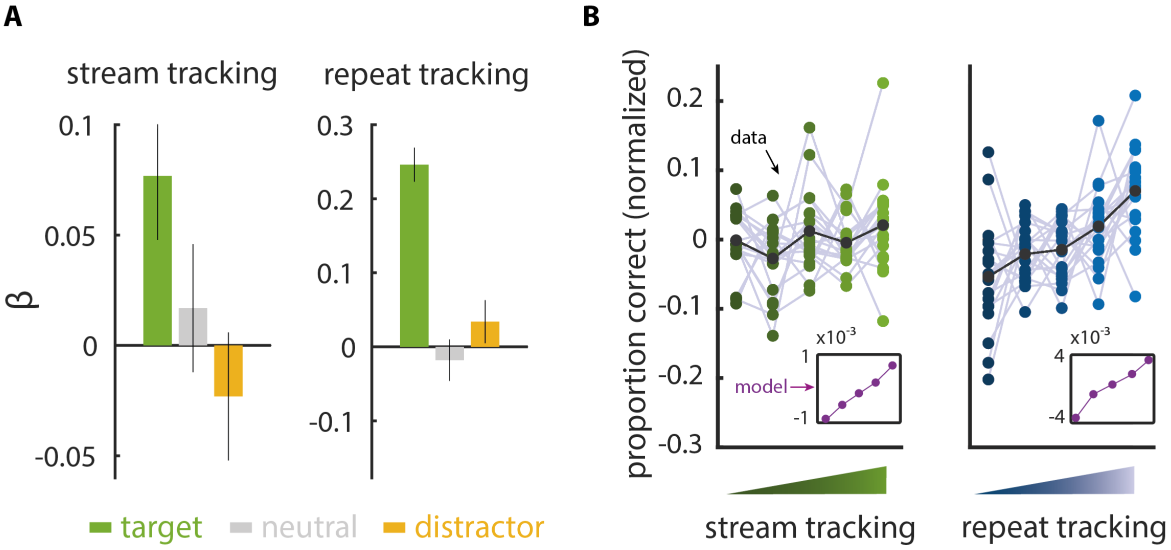


Figure 10: Brain–behaviour relation

Brain–behaviour relation. **A.** Standardized estimates (fixed effects, with SE) for the prediction of binary response behavior (hit vs. miss) by speech and repeat tracking for the target (green), neutral (gray) and distractor stream (orange).**B.** Coloured dots and gray lines show single subject proportion correct scores; black dots and a black line show the average across (N=19) subjects. For illustration, data were binned by stream/repeat tracking and normalization was done by subtracting the mean of single subject data across all bins from each corresponding subject data bin. Inset shows the model prediction for each bin.

Note especially that the estimates for the target stream and distractor stream pointed in opposite directions (Fig. 10A, left panel). We thus used a Wald statistic to test if the two estimates differed significantly from each other. The behaviour-beneficial contribution of the neural tracking of the target stream was positive and differed significantly from (as per sign of the estimator, behaviour-detrimental) neural tracking of the distractor stream (Z_{Wald} = 2.44, p = 0.015). As to be expected, the smaller differences between the neutral and target estimates (Z_{Wald} = –1.44, p = 0.147) and the neutral and distractor estimates (Z_{Wald} = 0.97, p = 0.332) proved not significant.

To control for potential confounding of the speech tracking in the target stream by the neural response to the to-be-attended repeats, we also included neural repeat tracking from all three streams in our model. Unsurprisingly, we observed a positive linear relationship between participant's performance and neural repeat tracking ($\beta = 0.246; SE = 0.023; z = 8.235; p < 0.001$) in the target stream. This shows that stronger neural responses to repeats in the target stream were associated with better behavioural detection of repeats. On the other hand, we observed no significant linear relationship between the tracking of a repeat in the neutral ($\beta = -0.018; SE = 0.028; z = -0.644; p = 0.520$) or in the distractor stream ($\beta = 0.034; SE = 0.029; z = 1.197; p = 0.231$; Fig. 10A, right panel). For illustration only, we binned the data by the strength of stream and repeat tracking into five bins (Fig 10B, right panel).

3.4.6 Control Analysis I: Listeners process the content of competing speech streams

The behavioural outcome from the comprehension questions was not of major interest to us, since the detection of repeats provides a much more reliable and finely resolved measure of behavioural performance. However, one concern we aimed to alleviate was that participants might have been only detecting repeats rather than listening to the speech content of the target stream at all; acoustic–phonologically processing the speech streams alone would probably be enough to identify the repeats. To further explore the degree to which listeners processed the speech streams semantically, 15 multiple-choice comprehension questions addressing all three streams were provided at the end of the study.

We used double iterative bootstrapping to estimate the 95% CI for the difference between the percentage of correctly answered questions and the previously determined empirical chance level of 40% (N=9 different participants only answering the questions without exposure to the full audio books; see Methods). By design, we were not able to differentiate between percentages of correctly answered questions in the target and distractor streams, as these switched their roles on a trial-by-trial basis. For instance, some questions required processing on a time scale that exceeded the trial length of 20s, which meant that some parts of the respective audiobook content belonged to the target and others to the distractor. Hence, we combined the correctly answered questions from the target and distractor streams (50±2%, mean±SEM, range: 30-67%).

This average response accuracy was significantly better than the empirical chance level (CI: 4.6–14.2% above chance). The percentage of correctly answered questions of the neutral audio stream was closer to chance (48±3%, mean±SEM, range: 27–80%), but there remained a significant if slim difference against the empirical chance level (CI: 0.9–14.6% above chance). The percentages of correctly answered questions did not differ systematically for the target/distractor stream versus the neutral stream (CI: –3 – 6.3%).

3.4.7 Control Analysis II: Condition-to-location assignment does not confound interference by distracting speech and sub-processes of attention

In a further control analysis, we considered the possibility that the spatial condition-to-location assignment could have an indirect effect on our behavioural and neural measures. Between subjects, we varied the position of the neutral sound stream (neutral: front/left/right). The different positions of the neutral stream lead to a different assignment of the target and distractor streams. The spatial separation between the target and distractor streams was 90° when neutral was presented at 0° and 45° when neutral was presented at 45° or -45°. To control for the different spatial condition-to-location assignments, we included the factor condition-to-location assignment as a covariate in our behavioural and neural analysis.

In our behavioural analysis, we observed a significant main effect of the factor condition-to-location assignment (F = 4.47; df = 15; p = 0.03). This effect is mostly driven by a significant difference between the condition-to-location assignment: neutral front versus neutral right (t = 2.96; df = 15; p = 0.01). In other words, participants correctly detected more repeats when the neutral stream was presented in front compared with the neutral stream presented on the left or right. There was no significant difference between neutral front versus neutral left (t = 1.29; df = 15; p = 0.22) and neutral right versus neutral left (t = -1.71; df = 15; p = 0.11). Importantly, however, the difference in sensitivity was independent of the spatial position of the neutral stream: There was no significant interaction between the factors attention and condition-to-location assignment (F = 1.44; df = 15; p = 0.268).

In our neural analysis, the main effect for the factor condition-to-location assignment was not significant (F = 0.328; df = 16; p = 0.725). Importantly, the differences in neural tracking were independent of the spatial position of the neutral streams. There was no significant interaction between the factors attention and condition-to-location assignment (F = 0.88; df = 32; p = 0.482). In sum, between-subject differences in the spatial condition-to-location assignment did not confound our results.

3.4.8 Control Analysis III: Unfolding of neural filters (TRFs) across trial duration

To account for the possibility that attentional processes such as enhancement, capture, and suppression unfold on different time scales over the trial duration and might cancel each other out, we divided the 20-s trial into 4, non-overlapping windows of 5 s each and estimated TRFs separately for each window (Fig. 11). Cluster permutation tests revealed that target enhancement is sustained across the trial duration. Importantly, we found no significant clusters for the distractor-vs.-neutral contrast (i.e., no evidence for capture or suppression). Also, a temporally more finely resolved analysis revealed no evidence for distractor capture or suppression. This analysis further supports our finding that target enhancement (i.e., attentional gain) is the dominant mechanism that modulates the neural phase-locked response to competing speech in a cocktail party scenario.

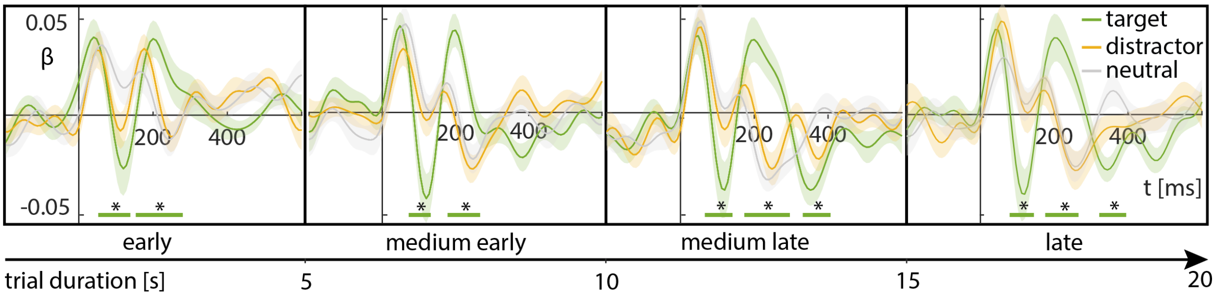


Figure 11: TRFs across trial duration

TRFs across trial duration. TRF β-weights are estimated in four separate 5 s time windows across the trial duration (20s), representing early to late attentional processing during the trial. TRF β-weights are averaged across subjects (N=19) and channels of interest: Fz, Cz, CPz and Pz (solid lines). Shaded areas show the standard error for each time lag across subjects. Cluster permutation test shows significant clusters between target and neutral speech tracking in each time window (green bars). No significant clusters for distractor versus neutral speech tracking are observed.

3.5 Discussion

The present study aimed to test whether the human auditory cortex enhances targets or suppresses distractors when implementing selective attention to continuous speech. To do so, we have proposed a new, three-stream continuous-speech design with an embedded psychophysical task. The most important results can be summarised as follows:

First, the paradigm is feasible to delineate different sub-processes of auditory attention, separating a task-relevant target speech stream better from potentially neutral speech than from distracting speech. This finding proved robust under analyses controlling for stream location relative to the listener.

Second, the neural results suggest that attention is implemented through enhancement of the target stream. This lack of neural differentiation of tracking a distracting vs. tracking a neutral stream speaks against mechanisms of ”active” or below-baseline neural suppression of distractors at the level of the human auditory cortex as measured with EEG.

Third, in line with an enhancing neural attention mechanism, the momentary neural tracking of the target but not the neural tracking of other, competing streams can predict the momentary likelihood that a listener detects events in this target stream.

3.5.1 Neural tracking of speech implements enhancement, not suppression

As in previous studies (e.g., Di Liberto et al., 2015; Ding & Simon, 2012; Fiedler et al., 2019; Har-shai Yahav & Zion Golumbic, 2021; Kerlin et al., 2010; Kraus et al., 2021; Lalor & Foxe, 2010) we found the strongest neural tracking for the target stream, which was mainly due to enhanced N1 and P2 components of the cortical response. Notably, this improved tracking could be due to increased sensory gain, but it could also be due to more precise temporal fidelity of the target stream, or both (Ponjavic-Conte, Hambrook, Pavlovic, & Tata, 2013). Critically extending these previous findings by implementing a neutral, task-irrelevant ”baseline stream” in a three-talker paradigm, we were able to assign these previous findings to two sub-processes of selective attention: target enhancement and distractor suppression. We found a significant difference in neural tracking between target and neutral streams but no significant difference between distractor and neutral streams.

We found that participants erroneously detected more repeats in distractor versus neutral speech, which indicates attentional capture on the behavioural level. Despite this signature of capture in behaviour, we found neither suppression nor capture in the neural speech tracking response. In the visual modality, it was shown that capture and suppression go together. A distractor can capture attention, followed by suppression thereafter (Gaspelin & Luck, 2018). We have addressed this issue by analysing different time windows along the trial. However, we found no evidence for distractor capture or suppression, analysing early and late time windows separately. But that does not mean that suppression is not implemented on the cortical level in general. For instance, modulation of alpha oscillatory power is a potential neural mechanism that might implement distractor suppression in a scenario with competing auditory streams (Wöstmann, Alavash, & Obleser, 2019).

Neural tracking of ignored speech is modulated by signal-to-noise ratio (SNR), hearing loss and perceptual demand. Fiedler et al. (2019) showed that SNR manipulations of ignored speech led to differential modulation of ignored speech and the resulting neural tracking. Also, hearing loss differentially affected neural tracking of attended versus ignored speech (Petersen et al., 2017). Recently, it was found that neural tracking of distracting speech in noisy auditory scenes depends on perceptual demand (Hausfeld, Shiell, Formisano, & Riecke, 2021). Following a rationale established before in visual neuroscience (Seidl et al., 2012), we manipulated the attentional fate of ignored speech by varying the listener's need to minimize or eliminate interference generated by the (previously task-relevant) distractors.

There is plenty of experimental evidence suggesting that selective attention is mainly enhancing the neural signal–to-noise ratio, thus effectively clearing or sharpening target representations in the visual and auditory domain (Desimone et al., 1995; J. B. Fritz, Elhilali, David, & Shamma, 2007; Gazzaley, Cooney, McEvoy, Knight, & D'esposito, 2005; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; McAdams & Maunsell, 1999; Mesgarani & Chang, 2012; Peelle et al., 2013; Golumbic et al., 2013). In line with these findings, we show that the prioritization of the neural representation of the target auditory input is mainly implemented by an enhancement of the target. In this respect, our results are also notably in line with a recent visual EEG study on attentional suppression by Gundlach et al. (2022). Also, another recent study investigated whether exogenous attention led to facilitation of attended information, suppressed unattended information, or both (Keefe & Störmer, 2021). Both studies found that attention rather operates on target enhancement than distractor suppression.

Generally, our study adds to the unsettled debate in attention research over neural implementations of suppression. Even before the present study, evidence in the literature for distractor suppression has been mixed, with some studies speaking to (Desimone et al., 1995; Schwartz & David, 2018; Seidl et al., 2012; Wöstmann, Alavash, & Obleser, 2019) and others speaking against distractor suppression (Gundlach et al., 2022; Keefe & Störmer, 2021; Noonan, Crittenden, Jensen, & Stokes, 2018).

Classical theories of attention permit some form of distractor suppression (Broadbent, 1958; Treisman, 1960), and there might well be distinct types of distractor suppression as endpoints to a continuum. Also, from a neurocognitive vantage point, distractor suppression does not need to be one single process and could rather be implemented via multiple neural mechanisms.

Firstly, suppression could be driven by the current intention of the observer extracting statistical regularities of certain features such as location of a distractor over time, enabling the brain to learn to produce suppression (Wang & Theeuwes, 2018b; Wöstmann et al., 2022). In the long term (duration of the experiment), participants could learn based on statistical regularities the location (same location of distractor stream) and the voice of the talker (same voice). Secondly, in the short term (every trial), participants are cued (current intention) to attend to one stream and to suppress the distractor (negative priming). In principle, our paradigm might initialise both of these types of distractor suppression. While it is debatable whether the effect of our negative priming manipulation persists over the whole trial duration (probably decreasing over time), learning and using statistical regularities of the distractor over time should persist in the long term of the experiment. However, we found no significantly suppressed neural tracking of the distractor vs. neutral stream, which suggests that the neural speech tracking response does not implement distractor suppression. Contrary to our hypothesis, results hinted rather at a potentially stronger tracking of the distractor compared to the neutral stream although this was not a statistically robust observation in the present data. For future studies, it is nevertheless important to consider such an attentional capture of the distractor stream (Gaspelin & Luck, 2019). In addition, participants could also have left some residual attention to the distractor stream in terms of divided attention between the currently relevant target stream and the previously relevant distractor stream, which led to the potentially stronger tracking of the distractor compared to the neutral stream (Miller, 1982). However, given the high hit rate for the target and the comparably low false alarm rate for the distractor stream, it appears rather unlikely that participants used divided attention as a strategy at least over the entire trial duration.

Secondly, distractor suppression can be generally divided into proactive (processing before the distractor appears) and reactive suppression (processing after the distractor has captured attention; Chelazzi, Marini, Pascucci, & Turatto, 2019; Wöstmann et al., 2022). The amplitude of neural alpha oscillations (10 Hz) related to top-down selective attention processes can be modulated by target- and distractor-processing. Wöstmann, Alavash, and Obleser (2019) found that alpha power during the anticipation of competing tone sequences implements distractor suppression independent of target enhancement. In a behavioural study, it was shown that the intelligibility of the target is improved when the masker is a familiar voice (Johnsrude et al., 2013). Their findings suggest that the brain uses a prior model of the characteristics of the distractor to actively suppress the distractor. In sum, the aforemetioned results speak to a proactive implementation of distractor suppression. But neural tracking is characterized by the time-lagged neural responses that phase-lock to the stimulus. Due to this characteristic, neural tacking is rather suited to investigate reactive suppression than proactive suppression. With respect to these distinguishable sub-processes of distractor suppression, our results indicate that at least reactive suppression is absent for auditory cortex responses in a multi-talker situation.

3.5.2 Auditory attention exploits statistical regularities to separate distracting versus neutral speech

When considering how distracting versus neutral, task-irrelevant speech might be encoded neurally, a previous auditory study using also three streams had suggested that higher-order auditory areas provide an object-based representation for the foreground, but the background remains unsegregated (Puvvada & Simon, 2017). At first glance, our results are broadly in line with this conclusion, but note that Puvvada and Simon had not applied any differential task manipulation to the two background speech streams, which we aimed to achieve here. The here proposed experimental paradigm aimed to strike important compromises in studying the listener's neurocognitive ability to separate target, distractor and neutral speech.

In contrast to trial-based designs, continuous speech paradigms often lack rich behavioural data. Usually, comprehension questions regarding the content of the audio streams are asked to differentiate between attended and ignored audio streams (Broderick, Anderson, Di Liberto, Crosse, & Lalor, 2018; Fiedler et al., 2019). Asking comprehension questions has some drawbacks. Comprehension questions usually refer to a comparable long-time range. This limits the number of questions and thus the number of behavioural data that can be extracted from the experiment. Further, in our paradigm participants had to switch their attention every 20s between two audio streams, which did not allow us to strictly assign the question to attended or ignored parts of the audio streams. Hence, it was insufficient to ask comprehension questions solely to investigate the listener's cognitive ability to separate target, distractor, and neutral speech on the behavioural level. More fine-grained behavioural data were needed, ideally without losing much of the ecological validity of natural speech.

We used short repeats in the audio streams to obtain rich behavioural data. In trial-based designs, participants are asked much more frequently to respond, which also ensures a steady engagement in the listening task. Marinato and Baldauf (2019) also embedded short repeats in auditory objects, arguing that such a detection task requires the processing of the acoustic stream at the level of auditory objects. Such a repeat detection task might thus be particularly suited to study object-based mechanisms of selective attention. Adopting this approach here, we found that participants detected much more repeats in the target (hits) compared to the neutral and ignored stream (false alarms).

Recall that, in our paradigm, participants had to switch attention between the same two streams while they had to ignore the never-task relevant neutral stream. Importantly, we found a significantly larger behavioural interference by distractor speech than by neutral speech, but what is the underlying mechanism? Our results suggest that the neural fate of a stream on the previous trial has the potency to make it more distracting and capture attention on the text trial. This corresponds with the concept of negative priming. Negative priming refers to the effect that the reaction to a stimulus that was previously ignored is more error-prone and slower (Tipper, 1985). Classical negative priming designs consist of two main components: the prime (trial N) and the probe (trial N+1). The prime presents a certain stimulus (or stimulus feature) as a distractor, which becomes the target in the probe trial. Negative priming has been studied in vision in a detailed manner (E. Fox, 1995; May, Kane, & Hasher, 1995).

Although there are fewer studies that investigated negative priming in auditory selective attention, they reported similar results (Frings et al., 2015). Nowadays, most researchers agree that auditory negative priming (similar in vision) is explained by inhibition and retrieval theories (Frings et al., 2015). Longer response times and higher error rates are typically observed relative to a no priming condition (Banks et al., 1995; Mayr, Buchner, Möller, & Hauke, 2011; Mayr & Buchner, 2007). Notably, we did not present the same segments of the audio streams on two consecutive trials. Participants had to attend and ignore different segments of the audio streams in each trial, due to the ongoing structure of continuous speech. We assume that it was rather the spatial location or/and the voice that was associated with negative priming and leaked into the present trial, than the identity of the auditory stimulus. On the one hand, if a listener attended to a specific feature of an auditory object, not only this specific feature is enhanced, but all features related to the selected object (for review see, B. G. Shinn-Cunningham, 2008). On the other hand, one could argue that this also holds for features concerning negative priming and object suppression.

A more recent study varied randomly the location of the target and distractor and the speaker (Eben et al., 2020). They demonstrated negative priming in auditory selective attention switching with the spoken material. In sum, our new paradigm has proven feasible to utilise the negative priming phenomenon to unravel listeners' separation of distractor speech versus neutral speech.

3.5.3 Neural tracking of target but not distractor explains performance

Continuous speech paradigms often lack rich behavioural data. But only if we unravel the precise relationship between brain and behaviour can we reach a veridical understanding of cognitive processes such as selective attention (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). We embedded short repeats into the speech streams, which served as a trial-by-trial measure for behaviour. In addition, this also enabled us to predict behaviour from neural responses on a single-trial level. We found that neural tracking of the target stream only predicted trial-by-trial variation in repeat detection. Our results not only provide support to the functional relevance of neural speech tracking (Tune et al., 2021), but significantly expand this by providing an explanation for the underlying sub-processes of auditory selective attention, that is, enhancement of the target and not suppression of distractors predicts performance. In addition, this finding supports the feasibility of our new continuous speech paradigm since we found a significant relation between the neural tracking of continuous speech and the repeat detection behaviour. Further, the finding supports our previous findings since only target enhancement predicts behaviour. Indicating that the prominent process of selective attention is target enhancement rather than distractor suppression.

3.6 Limitations

There are limitations regarding the operationalization of the neutral and distractor streams. First, the attentional manipulation by their respective task-relevance (Seidl et al., 2012) of the distractor stream might not lead to an interference strong enough that distractor suppression was useful. Thus, it is possible that negative priming in combination with the spatial and/or spectral separation of the audio streams was insufficient to activate the need of distractor suppression in our study. Future studies could address this by varying for instance, the separation between the audio streams (Hausfeld et al., 2021). The task may become more difficult with smaller spatial separation, which potentially activate distractor suppression.

In addition, our sample size (N = 19) could have been too small to detect small distractor suppression effects. Note, however, that any such distractor-suppression effect size would need to be put in perspective given the considerable effect sizes of target enhancement we observed. So, the relative conclusion about target enhancement vs. distractor suppression would remain. Thus, the conclusion stands that target enhancement is the behaviourally and neurally more prominent sub-process of selective attention in a continuous speech paradigm.

3.7 Conclusion

In attention research, previous paradigms have rarely aimed at conclusively separating mechanisms of distractor suppression from mechanisms of target enhancement. Using a new, psychophysically augmented continuous-speech paradigm with three speech streams, our results demonstrate that neural tracking of continuous speech reflects target enhancement, not distractor suppression. These findings call for a refinement of current models about enhanced neural responses to speech and should account for specific sub-processes of selective attention, that is, the enhancement of targets rather than the suppression of distraction.

4 Exploring the interplay between dynamic range compression and selective attention in competing-talker environments

4.1 Introduction

In everyday life, people often encounter challenging hearing situations where multiple auditory signals are present. Selective attention allows listeners to prioritise a target auditory signal over distracting signals that may be occurring simultaneously (Desimone et al., 1995). People with normal hearing are remarkably adept at focusing on relevant signals (even complex signals like speech) while filtering out concurrent distractions (Cherry, 1953). However, individuals even with mild to moderate hearing impairments often struggle in multi-talker situations. Hearing aids are the most common treatment for people suffering from hearing impairment. In our study, we explored the impact of amplitude compression on the ability to focus attention and how it interacts with attention. We hypothesised that by compressing the amplitude of ignored talkers, the distinction between attended and ignored talkers would be enhanced, resulting in improved performance for the listener.

In recent years, computational techniques have been developed to estimate neural responses to single continuous auditory stimuli, even in the presence of other sounds (Crosse et al., 2016). Electrophysiological responses in cortical regions phase-lock to speech features in magneto/electroencephalogram recordings (Luo & Poeppel, 2007). The ”temporal response function” (TRF) captures this linear relationship between continuous speech features and neural response and can be interpreted in close analogy to the classical ERP (Crosse et al., 2016; Fiedler et al., 2019). Neural phase locking to the low-frequency envelope of speech, referred to as ”neural speech tracking” (Obleser & Kayser, 2019), serves as an objective measure for differentiating attended speech from concurrently ignored speech. Numerous studies have shown that individuals with normal hearing exhibit stronger neural phase locking to the envelope of attended speech compared to ignored speech (e.g., Brodbeck & Simon, 2020; Ding & Simon, 2012; Fiedler et al., 2019; Puvvada & Simon, 2017; Golumbic et al., 2013). Additionally, there is evidence that neural phase locking to the envelope of speech correlates with speech intelligibility (Peelle et al., 2013), as well as behavioural indices of speech comprehension (Etard & Reichenbach, 2019), and that stronger speech tracking enhances trial-to-trial behavioural performance (Tune et al., 2021).

Since neural tracking can be an objective measure for selective attention and correlates with behavioural measures, it is an interesting basis for research concerning the hearing-impaired system. However, the literature provides mixed evidence on how hearing impairment affects the neural tracking of the speech envelope. Early studies showed that poorer hearing was related to stronger tracking of the ignored envelope (Petersen et al., 2017). On the other hand, more recent studies suggest that hearing-impaired listeners show stronger neural tracking compared to the age-matched control group (Fuglsang et al., 2020). In contrast, other studies found no differences between older listeners with normal and impaired hearing in neural speech tracking (Goossens, Vercammen, Wouters, & van Wieringen, 2019; Presacco, Simon, & Anderson, 2019). The contradictory effects of hearing loss on neural tracking may be due to the complex interplay between ageing, the severity of hearing loss, and cognitive abilities. More recently, Schmitt et al. (2022) reported enhanced speech tracking with increasing hearing loss and suggested that the hearing impaired rely more on the tracking of slow modulations in the speech signal to compensate for their hearing deficit.

It's worth noting that changes in neural tracking have been observed in studies related to auditory processing. For instance, vocoding can lead to delayed neural separation of competing speech during attentional selection (Kraus et al., 2021), and late cortical tracking of ignored speech is modulated differently based on signal-to-noise ratios (Fiedler et al., 2019). Furthermore, a recent study found that neural speech tracking can serve as an indicator of the benefits of hearing aid algorithms, including amplitude compression (Petersen, 2022). Overall, these findings suggest that neural speech tracking could be a useful tool for researchers seeking to understand the effects of various hearing aid algorithms, such as dynamic range compression.

Dynamic range compression is an audio signal processing algorithm that amplifies quiet sounds while reducing the intensity of loud sounds. Dynamic range compression is commonly used in hearing aids to compensate for loudness recruitment in hearing impaired listeners with presbyacusis and to restore the outer world audio dynamic into the listener's hearing range (Kates, 2005). However, dynamic range compression also leads to undesired side effects. For instance, compression directly affects the envelope of a speech signal. It reduces the amplitude modulation depth, impairs the envelope shape, and leads to abrupt changes in the onsets and offsets of the speech envelope (Stone & Moore, 1992). Since the envelope of speech is not just a simple acoustic feature of speech but also associated with speech comprehension (for review, see Poeppel & Assaneo, 2020). We assume that dynamic range compression impairs, in general, the neural tracking of speech. However, hearing aids are able to perform spatial signal processing. This allows hearing aids to apply different compression ratios to different spatial locations of signals. If a comparable strong compression is only applied to ignored speech, this could be a potentially useful tool in a multi-talker situation. In addition to a pure reduction of the SNR, which has the consequence that ignored speech (noise) is very strongly down-regulated and makes a change of attention more difficult. In contrast, a higher compression ratio for ignored speech may provide a sweet spot between suppression and the ability to switch speakers situationally. We hypothesise that amplitude compression on the ignored stream increases the behavioural and neural separation between the attended and ignored streams, as reflected by faster response times and increased behavioural responses. This hypothesis is based on the assumption that compression on ignored talkers will reduce their salience and facilitate their suppression by attention, thereby leading to improved performance on the attended task.

To test our hypothesis, we first conducted a pilot experiment to determine the appropriate compression ratio for the following main experiments. We then recruited 24 normal hearing control participants to participate in a quasi-categorical paradigm in which the speech streams could be unprocessed, compressed, or one of the two streams could be compressed while the other was not. Following this, we modelled the human auditory periphery using the model from (Verhulst et al., 2018) to investigate the peripheral fate of compressed and unprocessed speech in normal hearing and hearing-impaired participants. In the following, we measured hearing-impaired participants with presbycusis in the same experiment as normal hearing participants, accounting for overall sound pressure level. Finally, we conducted a behavioural control online experiment to verify our used loudness matching procedure.

4.2 Methods

4.2.1 Normal hearing participants

The participants in the current study were 24 young adults (18 female and 6 male), aged 18 to 34 (mean: 25.5). Each participant reported having a native language of German, having normal hearing, and having no prior neurological conditions. We measured pure tone audiometry between 250 and 4000 Hz to confirm normal hearing. For the tested frequencies, all participants displayed auditory thresholds below 20 dB. They provided written, fully informed consent and were paid 10 euros per hour. The study was approved by the local ethics committee of the University of Lübeck.

4.2.2 Hearing impaired participants

We recruited a total of 10 participants with hearing loss, but three of them had to be excluded from the analysis. The first participant was excluded due to data loss during EEG recording, the second participant was excluded because they did not perform the task correctly, and the third participant was excluded because their hearing loss was due to acute causes in the right ear, rather than presbycusis. The study enrolled individuals between the ages of 50 and 75 with mild to moderate presbycusis, defined as a pure tone average (PTA) between 20-50 dB HL (Humes et al., 2012), similar hearing thresholds in both ears with a maximum difference of 10 dB, and no or less previous experience with compression in hearing aids, either unaided or with no longer than one year of prior use. The remaining N = 7 participants were aged between 60 to 73 years old, with an average age of 68.4 years. This also means that both groups were not matched in age. To assess their hearing ability, pure tone audiometry was performed for frequencies ranging from 250 to 4000 Hz. All participants showed the typical sloping progression to higher frequencies in auditory thresholds, with pure tone averages ranging from 26-40 dB HL and an average of 33.1 dB HL. Written, fully informed consent was obtained from all participants, and they were compensated at a rate of 10 euros per hour. The study was approved by the local ethics committee of the University of Lübeck.

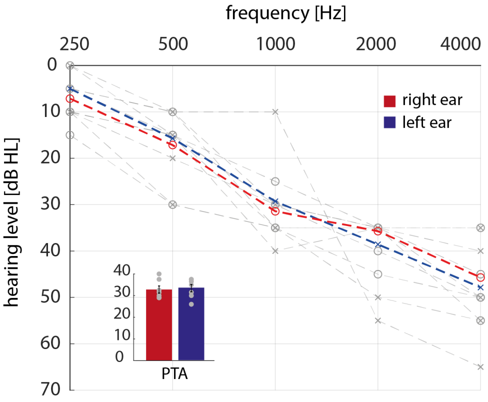


Figure 12: Audiogram and PTA

4.2.3 Stimulus materials

We presented audio versions of two different narrated book texts, ”Ludwig van Beethoven Basiswissen” and ”Sophie und Hans Scholl Basiswissen”, both of which were spoken by professional talkers. We selected audio streams that had not previously undergone amplitude compression. At an average intensity (SPL mixture) of about 65 dB(A), which is comparable to the volume of a normal conversation, the two audio streams overlapped in time.

Using customised MATLAB code, the stimuli were processed in the following steps (Version 2018a Mathworks Inc., Natick, MA, United States). The audio files had a 44.1 kHz sampling rate and a 16-bit resolution. The maximum duration of silent periods was reduced to 500 ms (O’Sullivan et al., 2015).

By selecting 400 ms of the original audio stream and repeating it immediately after, we added brief repeats to both audio streams (Marinato & Baldauf, 2019). At least two seconds after the stimulus began, the first repeat was shown. By linear ramping and cross-fading, each repeat was incorporated into the sound stream. Utilizing a window of 220 samples (5 ms) from the down ramp’s end and the first 220 samples (5 ms) from the repeat itself, linear ramping was performed (up ramp). The cross-fading was accomplished by combining the up and down ramps.

In order to prevent undetectable repeats of weak sound intensity, we further used an rms (root mean square) criterion, which required that the repeat’s rms be at least equal to the rms of the stream from which it was drawn.

Using a digital dynamic range compressor built into MATLAB, we applied amplitude compression to the speech streams (Giannoulis, Massberg, & Reiss, 2012). The following is how we set the compressor parameter: Attack time: 2 ms; release time: 15 ms; threshold: -40 dB. We used a high compression ratio and relatively quick attack and release times compared to standard hearing aid processing (e.g., Kates, 2010). Then, to avoid clipping, the uncompressed and compressed audio books were both limited to 99.95% (Figure 13). The compression ratio was determined by the pilot experiment (Section 4.3).

We created four pairs of segments -uncompressed & uncompressed, uncompressed & compressed, compressed & uncompressed, and compressed & compressed- each belonging to both of the streams that were simultaneously presented in order to maintain a balance between compressed and uncompressed segments in the two streams. Each pair had a duration of 5-minutes. The pairings were arranged in a balanced and random manner.

The root-mean-square (rms) is frequently used to match the intensity of distinct audiobooks. However, it was demonstrated that the perceived loudness of RMS matched unprocessed and compressed speech differs (Moore, Glasberg, & Stone, 2003) Here, we matched the perceived loudness for time-varying acoustic signals based on Zwicker using an internal MATLAB function (Zwicker & Scharf, 1965). We conducted a psychophysical experiment to confirm the viability of this algorithm using the stimuli we had previously used (Section 4.7).

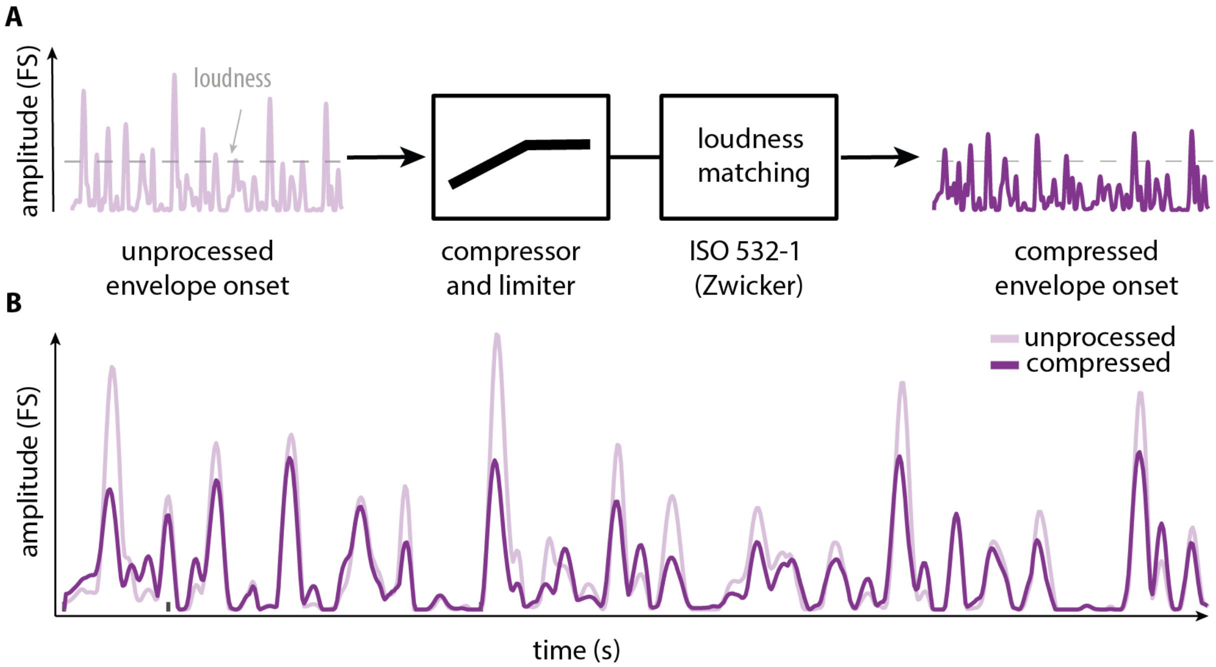


Figure 13: Stimulus processing pipeline

Stimulus processing pipeline. The most important stimulus processing flow is shown by **A**. The stimulus is expressed as the envelope onset of the speech signal. First, the signal is processed by the compressor (ratio: 1/8, attack time: 2 ms; release time: 15 ms; threshold: -40 dB). Importantly, the limiter was then applied to both signals the uncompressed and compressed signals- to avoid clipping. The compressed audio segments were matched to the loudness of their uncompressed counterparts by using a MATLAB algorithm based on Zwicker and Scharf (1965). The same stimulus envelope onset is shown uncompressed and compressed with loudness matching in **B**.

The cue was presented at the center of the screen (resolution: 1920x1080, Portable HDMI Screen, Wimaxit) in front of the participant (distance: 1 m). The spatial cue consisted of two sub-triangles which had a size of 1.3° visual angle pointing to the front and back sound sources The two triangles had different colours blue and red. Participants had to attend either to the red or the blue triangle. Since the cue and the fixation cross were presented at the same time as the auditory stimuli, we ensured that the possible interference between visual and auditory neural responses was as small as possible. In order to achieve this, the cue was linearly faded in and out (50 ms each) to create a seamless transition between the fixation cross and cue.

4.2.4 Experimental setup

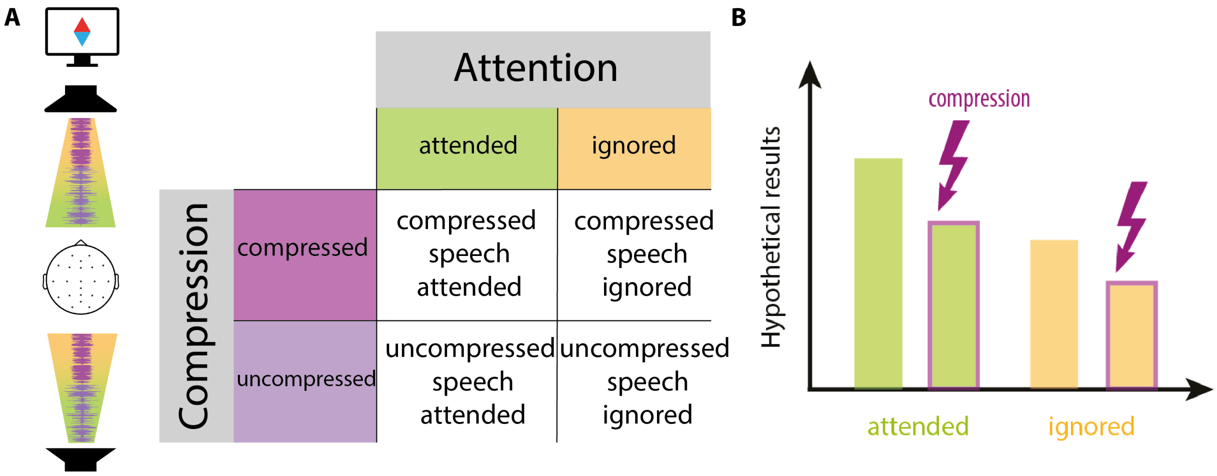
The experiment was conducted in a soundproof chamber with two loudspeakers (Genelec: Speaker 8020D, Denmark) placed at a one-meter radius in the front and back. The ground was 1.20 meters away from the loudspeakers, and a chair was positioned in the center of the radial speaker array with its face aligned to the loudspeaker at position 0° in the azimuth plane. Participants received a briefing on the experiment in advance. Each participant was instructed to keep their eyes open, keep their gaze on the center of the screen, and sit as comfortably as they could. A chin rest was utilized to prevent head movement. Each participant had their chin rest adjusted in height.

4.2.5 Experimental procedure

To study amplitude compression in a multi-talker paradigm, we developed a new experimental procedure. The experiment was created using Psychophysics Toolbox extensions (Brainard & Vision, 1997; Pelli & Vision, 1997) and MATLAB (MathWorks, Natick, MA, USA). Two audio streams were played simultaneously for participants. Each trial started with a cue that specified which stream to attend, displayed for 500 milliseconds. A fixation cross was then shown for the remainder of the trial (19.5 s), while the auditory stimuli continued to play in the background. Trials were presented continuously, with the next trial starting immediately after the previous one ended.

Participants were required to identify short repeats in the target stream, with six repeats included in each trial and randomly divided between the two streams. Prior to data collection, the experiment was explained to participants, emphasising the importance of listening to the target stream's content and responding as quickly and accurately as possible to a repeat. Participants were given a single sentence containing one repeat to acquaint them with the repeats, and were asked to provide oral feedback if they were able to recognise it. Additionally, participants completed six practice trials that were identical to the main experiment but used different audio streams. The main experiment lasted approximately one hour and included 196 consecutive trials split into four blocks, with participants having the opportunity to rest between each block.

At the conclusion of the experiment, we gave the participants 15 multiple-choice questions (each with four possible answers). We did not pose the questions after every block to prevent participants from paying attention to the audio stream that was meant to be ignored. Participants were presented with the questions in a certain order and a set of potential answers that were randomised.



		Attention	
		attended	ignored
Compression	compressed	compressed speech attended	compressed speech ignored
	uncompressed	uncompressed speech attended	uncompressed speech ignored

Figure 14: Experimental paradigm

Experimental paradigm and hypothetical results. **A** Left: Experimental setup. Two loudspeakers are placed in front (0°) and back (180°) of the participant. Speech streams are simultaneously presented over both loudspeakers. A screen was placed in front of the participant. A spatial cue indicated to which location participant had to attend. Right: Experimental paradigm. We had a quasi-factorial design with the factors attention (2-levels:attended and ignored) and compression (2-levels: compressed and uncompressed). Importantly, attending and ignoring always happened at the same time, while the factor compression was fully balanced. **B** Hypothetical results. For the main effects, we would expect that attention (to-be-attend to the cued stream) has a positive effect on the dependent variables, which means that attention leads to increased behavioural and neural results. In contrast, compression would have a negative effect on the dependent variables, that is, decreased neural and behavioural results. More importantly, we would expect an interplay between attention and compression. Compression on the ignored stream, in particular, increases neural and behavioural separation when compared to when no compression is applied to the ignored stream.

4.2.6 Sound pressure level adjustment for hearing impaired participants

The hearing-impaired participants completed the same experiment as their normal hearing counterparts, using the same stimuli but with different randomizations of conditions. However, we made one change: We adjusted the overall sound pressure level of the experiment based on each participant's hearing loss. To determine their hearing threshold, we used 500 ms parts of the stimuli that were presented in the experiment itself over free-field loudspeakers. We employed a combination of limit and constant stimuli methods to establish the threshold for the experimental stimuli. First, we presented pairs of compressed and uncompressed stimuli snippets (one over the front loudspeaker, one over the back), with one always louder than the other. In 3 dB steps, we decreased the sound pressure level of the signals each time the participant pressed a button to indicate they could hear the sound snippet. Once the participant stopped responding, we set this level as a reference for the method of constant stimuli. We then presented three different levels in 2 dB steps before and after the reference level, with each level presented 10 times in random order for a total of 70 presentations. We used the participants' responses to fit a psychometric function and obtained the SRT50 of this function as the new determined threshold. We added 35 dB to this threshold to set the presentation level. However, we asked each participant after the procedure if the overall presentation level was appropriate for them. If they did not agree, we adjusted the presentation level in 5 dB steps until it matched their reported most comfortable perceived loudness. On average, the presentation level was 72 dB ranging from 65 to 88 dB SPL.

4.2.7 Data acquisition and pre-processing

A 24 electrode EEG-cap (Easycap, Herrsching, Germany; Ag-AgCl electrodes positioned in accordance with the 10-20 International System) connected to a SMARTING amp was used to record the EEG (mBrainTrain, Belgrade, Serbia). The portable EEG system sends the signal via Bluetooth to a computer for recording . Using the program Smarting Streamer (mBrainTrain, version: 3.4.2), EEG activity was captured at a sample rate of 500 Hz. Impedances were kept under 20 kΩ while impedances were used as an online reference during recording using electrode FCz.

The Fieldtrip-toolbox, built-in functions, and MATLAB (Version 2018a Mathworks Inc., Natick, MA, United States) were used for offline EEG preprocessing (Oostenveld et al., 2011). High- and low-pass filters were applied to the EEG data between 1 and 100 Hz, and the electrodes M1 and M2 (the left and right mastoids) were averaged (two-pass Hamming window, FIR). On the EEG data from every participant, an independent component analysis (ICA) was performed. Prior to ICA, M1 and M2 were removed. Visual inspection was used to identify and remove ICA components associated with eye blinks, eye movement, muscle noise, channel noise, and line noise. On average, 7.89 out of 22 components (SD = 2.74), were disqualified. Back projected to the data were elements not connected to artifacts. Clean EEG data were processed further. Frequencies up to 8 Hz are associated with neural speech tracking (Luo & Poeppel, 2007). EEG data were therefore low-pass filtered once more at 10 Hz (two-pass Hamming window, FIR). EEG data were then segmented into epochs that matched the trial length of 20s and resampled to 125 Hz.

4.2.8 Extracting the speech envelope

By calculating the onset envelope of each audio stream, the temporal fluctuations of speech were quantified (Fiedler et al., 2017). In the beginning, we used the NSL toolbox (Chi et al., 2005) to compute an auditory spectrogram (128 sub-band envelopes logarithmically spaced between 90 and 4000 Hz). In order to create a broad temporal envelope, the auditory spectrogram was secondly averaged across frequencies. Third, the half-wave rectified first derivative of the onset envelope was obtained by computing the first derivative of this envelope and zeroing negative values. In order to match the EEG analysis's target sampling rate, the onset envelope was lastly down sampled (125 Hz). By using the onset envelope instead of the envelope, the envelope is moved in time. It's significant that the TRF obtained by using the onset envelope as a regressor resembles a conventional ERP the most (Fiedler et al., 2017).

4.2.9 Temporal response function and neural tracking estimation

A temporal response function (TRF) is a condensed brain model that illustrates how the brain would process the acoustic speech envelope of the stimulus to produce the recorded EEG signal if it were a linear filter. To calculate the TRF, we employed a multiple linear regression method (Crosse et al., 2016). In order to more precisely predict the recorded EEG response, we trained a forward model using the onset envelopes of the attended and ignored streams (e.g., Fiedler et al., 2019). In this framework, we examined delays between envelope changes and brain responses of between -100 and +500 ms.

To address EEG variance related to processing behaviorally relevant repeats and corresponding evoked responses, we added all repeat onsets and button presses as nuisance regressors using stick functions. These repeat onsets were added independently of the speech envelope regressors and chosen almost randomly (within SNR threshold) for each speech stream.

To prevent overfitting, we used ridge regression to estimate the TRF and determined the optimal ridge parameter through leave-one-out cross-validation for each participant. We predefined a range of ridge values, calculated a separate model for each value, and averaged over trials to predict the neural response for each test set. The ridge parameter with the lowest mean squared error (MSE) was selected as the optimal value specific to each subject. TRFs were estimated from trials in the experiment. To avoid cue conflicts, the first second of each trial was excluded. One model was trained on 192 trials using predictor variables for the onset envelopes of attended and ignored streams, as well as stick functions for repeats and button presses. These were modeled jointly (same regressor matrix) using the same regularization.

Neural tracking measures the representation of a single stream in the EEG signal, using TRFs to predict the EEG response. By using Pearson correlation to compare the predicted and actual EEG responses, the neural tracking (r) was calculated. By using the leave-one-out cross-validation method, we were able to predict the EEG signal on single trials (see above). A sliding-time window (48ms, 6 samples, 24ms overlap) calculated neural tracking accuracy over TRF time lags, resulting in a time-resolved neural tracking (Fiedler et al., 2019; Hausfeld et al., 2018; Kraus et al., 2021; O'Sullivan et al., 2015).

4.2.10 Statistical analysis

We employed various statistical methods to address our research questions. To examine the behavioural data in relation to detected repeats, we utilised logistic regression to model the binary outcome (hit = 1/miss = 0) of each repeat. We used a mixed model to predict the continuous dependent variable response speed. We incorporated both attention and compression categorical predictors in both models to examine their main effects and interactions. To investigate the attentional pairs (separate model), we incorporated a categorical predictor that indicated the corresponding information of the pair in which the dependent variable was measured. To assess statistical differences in neural tracking, we employed mixed models with the same categorical predictors as previously mentioned. However, the difference was that we utilised the models to predict neural tracking. Additionally, we included the categorical predictor space (front/back) to control the spatial assignment of loudspeakers in the setup for all models. We used jamovi for, gamlj package in R for fitting generalized linear mixed models (Jamovi Project, 2020), and MATLAB's fitlme function for fitting linear mixed models (MathWorks Inc., 2020).

4.2.11 Statistical analysis on time series

We investigated whether there were differences in time points in time-resolved neural tracking and TRF between subjects in different conditions and attentional pairs. To do this, we utilized a two-level statistical analysis known as cluster permutation test, which was implemented in Fieldtrip (Oostenveld et al., 2011). The analysis was conducted on data from 22 channels. At the single-subject level, we performed one sample t-tests to assess TRF and time-resolved neural tracking differences. Clusters were defined based on resulting t-values and a threshold set at $p < 0.05$ for at least three neighboring electrodes. The observed clusters were compared to 5000 randomly generated clusters through a permutation distribution using the Monte Carlo method to correct for multiple comparisons (Maris & Oostenveld, 2007). The cluster p-value was determined by the relative number of Monte Carlo iterations in which the summed t-statistic of the observed cluster exceeded that of the random clusters.

Due to the small sample size of only seven participants with hearing impairment and six participants for the pilot study, we employed double iterative bootstrapping to assess the statistical significance of our results. This method is particularly useful for small sample sizes and can help to mitigate issues with low power and unreliable results. Double iterative bootstrapping involves resampling both the participants and the data, and then running the analysis multiple times to generate a distribution of results that can be used to estimate the true population parameters. Instead of using a single bootstrap, we opted for a double bootstrap approach (J. Fox, 2016) to improve the accuracy of the confidence intervals. We performed the bootstrap procedure using the iboot package, which is designed for iterated bootstrap analysis of small samples and samples with complex dependence structures (Penn, 2020). We implemented the procedure in Matlab (MathWorks Inc., 2020) and used 2000 bootstrap samples and 200 repetitions for the inner loop based on the methods of Tibshirani (1993) and J. Fox (2016).

4.3 Study 2: The effects of compression ratio on neural tracking of speech: A pilot study

In this pilot experiment, seven young individuals with normal hearing were asked to listen to a narrative story that contained randomly balanced parts of three different compression and expansion ratios. The compression ratios were 1:2 and 1:8, while the expansion ratio was 2:1. The participants were also presented with a baseline condition where the narrative story was unprocessed. The task was to listen to the content of the presented narrative story. The objective of this pilot experiment was to identify an appropriate compression ratio to be used in a follow-up study.

For this pilot study, we opted to use a decoding approach rather than an encoding approach. The main reason for this choice was to take advantage of the higher accuracy that can be achieved by using all EEG channels for reconstruction also given the small sample. Additionally, we were able to avoid potential confounds related to the temporal response functions of the brain to repeated stimuli, as well as confounds related to button presses since repeats were not included in this pilot experiment.

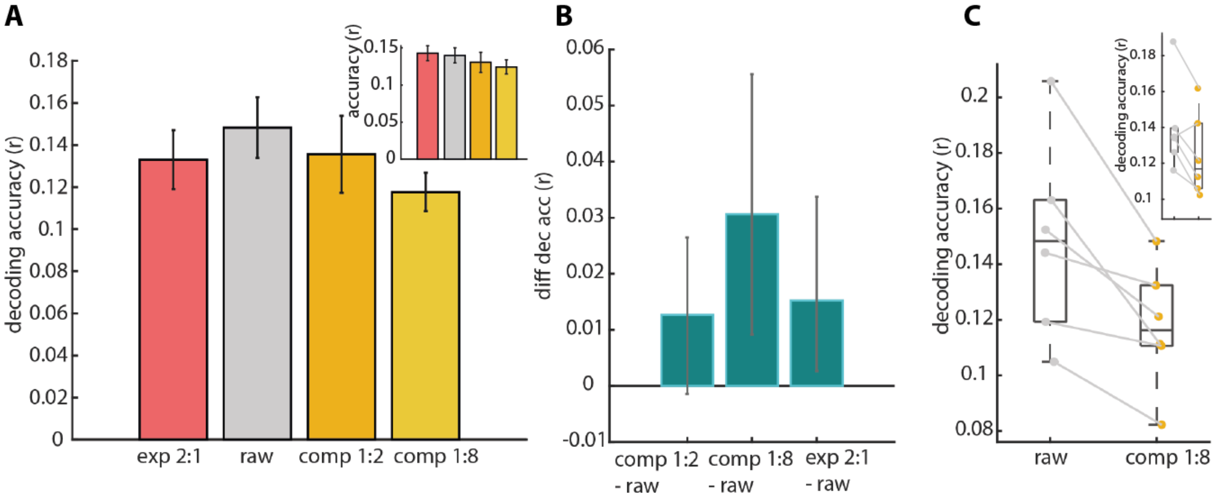


Figure 15: Pilot study

Neural results for different compression and expansion ratios. A Decoding accuracy refers to the Pearson correlation between the stimulus onset envelope and the estimated onset envelope, using all EEG channels. Inset: shows the decoding accuracy obtained by using the uncompressed onset envelops for all conditions. Bar plots indicate the participants' means, and error bars indicate the standard error of the mean (SEM). **B** Absolute difference in decoding accuracy between compression and expansion ratios and the unprocessed baseline (raw). Bar plots indicate the participants' means, error bars to indicate the 95% bootstrapped confidence intervals (CIs) against zero. **C** This plot shows the decoding accuracy for the unprocessed baseline compared to the 1:8 compression ratio, with dots representing individual subject data and connection lines indicating the same subjects.Inset: shows the decoding accuracy obtained by using the uncompressed onset envelops for all conditions. The boxes represent the interquartile range (25th to 75th percentile) of the data, with the median indicated by a line inside the box. The whiskers extend to the most extreme data point, excluding outliers.

Figure 15 A depict the averaged decoding accuracy for different compression and expansion ratios. To investigate differences between conditions we used bootstrapped CIs. For the comparison of compression ratio 1:2 and expansion ratio 2:1 versus unprocessed baseline (raw) the bootstrapped CI included zero, indicating a non-significant difference (Figure 15 B). However, for the difference between compression ratio 1:8 and raw the bootstrapped CI is not including zero, indicating a significant difference. That is, the decoding accuracy for the 1:8 compression ratio was significantly lower than that for the unprocessed speech, as indicated by a bootstrap CI that did not include zero. Interestingly, each participant shows a decreased decoding accuracy for the 1:8 compressed speech signal (Figure 15 C).

In all of our analysis, we used the onset envelopes of the actually presented signals. As a control, we also conducted an analysis using the uncompressed onset envelope, even when the signal was expanded or compressed, while keeping the ridge regression parameter λ constant (Figure 15 A Inset). In this alternative analysis, we observed that most (5 out of 6) participants demonstrated a decrease in neural tracking, as measured by decoding accuracy, for compressed (1:8) speech (Figure 15 C Inset).

Based on the results of the pilot experiment, we concluded that using a 1:8 compression ratio for amplitude compression reduces the brain's ability to track speech. Therefore, we chose to use this compression ratio in our main experiment, which also included attention as an experimental factor. We made this decision for two main reasons. Firstly, our initial hypothesis that compression on ignored streams would increase neural separation only works when participants focus their attention on one stream and ignore the other. Secondly, attention could also have affected the compression effect in the pilot experiment, as participants may not have been attending to the compressed (1:8) speech stream.

4.4 Study 3: The interplay between dynamic range compression and selective attention in normal hearing listeners

24 young, hearing-normal participants were simultaneously presented with two streams of continuously narrated audio. The assignment of amplitude compression to the speech streams was counterbalanced within participants. They had to alternate between the two audio streams on a trial-by-trial basis. Participants had to quickly and accurately identify any repetitions in the attended stream while ignoring the unattended stream.

4.4.1 Compression on both streams impairs performance

Here, we analysed the behavioural data in terms of the proportion of detected repeats and response speed (dependent variables). We tested the effects of attention and compression (independent variables) on the dependent variables. For the main effects, we would expect that attention (to-be-attend to the cued stream) has a positive effect on the dependent variables, which means that attention leads to an increased proportion of detected repeats and an increased response speed (inverse of response time). In contrast, compression would have a negative effect on the dependent variables, that is, a decreased proportion of detected repeats and a decreased response speed. We would expect a quasi-interaction between attention and compression. However, it is important to note that we do not have a classic factorial design but that the attended and ignored streams are always presented simultaneously, hence we contrasted attentional pairs against each other. Specifically, we would expect that the effect of attention depends on compression, with compression on the ignored stream increasing behavioral performance in contrast to the attentional pair where no compression is applied to the ignored stream.

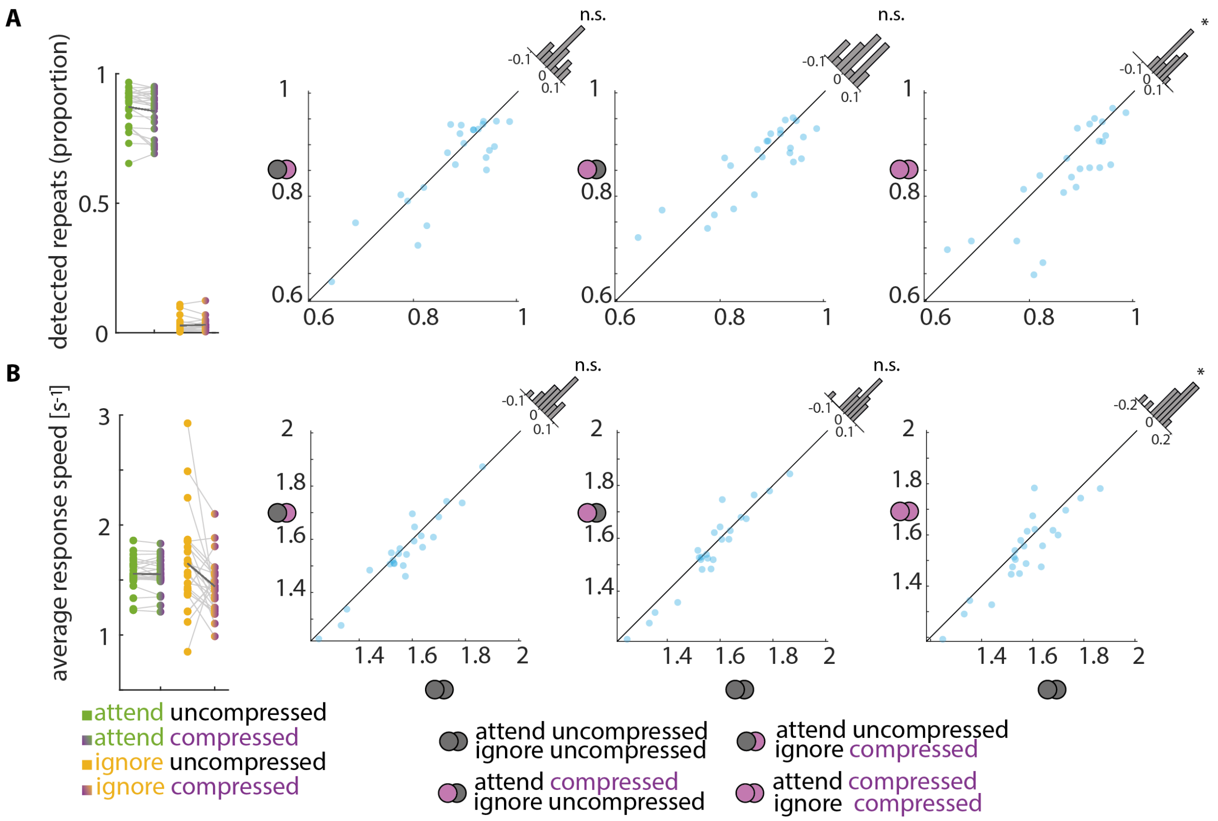


Figure 16: NH: Behavioural results

Proportion of detected repeats and corresponding response speed. **A** Left: Proportion of detected repeats per condition. Dots show individual (N=24) mean proportions of detected repeats. Light gray lines indicate the same subject. The bold line represents the mean of the group. Right: Proportion of correctly detected repeats in the attend stream displayed for attentional pairs (attend/ignore) with the four experimental compression combinations. The attentional pair (attend-uncompressed:ignore-uncompressed) served as a baseline and is contrasted against the remaining three pairs. Dots depict single subject data. Histograms show the distribution of the difference in correctly detected repeats for the contrasted attentional pairs. **B** Left: Response speed per condition. Dots show individual (N=24) mean response speed. Light gray lines indicate the same subject. The bold line represents the mean of the group. Right: Response speed of correctly detected repeats in the attend stream displayed for attentional pairs (attend/ignore) with the four experimental compression combinations. The attentional pair (attend-uncompressed:ignore-uncompressed) served as a baseline and is contrasted against the remaining three pairs. Dots depict single subject data. Histograms show the distribution of the difference in response speed of correctly detected repeats for the contrasted attentional pairs.

Participants were well able to detect repeats in the attended uncompressed stream (Figure 16 A left; mean accuracy: 0.87, 95% CI: [0.84, 0.90]; mean speed: 1.56 s^{-1}, 95% CI: [1.49, 1.62 s^{-1}]). They were also well able to detect repeats in attended compressed stream (mean accuracy: 0.86, 95% CI: [0.82, 0.89]; mean speed: 1.55 s^{-1}, 95% CI: [1.49, 1.62 s^{-1}]). They made a only few false alarms for the ignored uncompressed stream (false alarm rate: 0.03, 95% CI: [0.02, 0.04]; mean speed: 1.65 s^{-1}, 95% CI: [1.46, 1.84 s^{-1}) and for the ignored compressed stream (false alarm rate: 0.03, 95% CI: [0.02, 0.04]; mean speed: 1.44 s^{-1}, 95% CI: [1.34, 1.55 s^{-1}). Jointly, the number of hits and false alarms indicate that participants were attending to the cued speech stream. We found no significant difference in mean accuracy (b = 0.02, SE = 0.06, OR = 1.02, 95% CI [0.99, 1.14], p = .74) between the compressed and uncompressed streams.

Participants responded with similar response speed to repeats in attend and ignored speech, no significant differences were observed (Figure 16 B left; b = 0.036, SE = 0.05, t(12268) = 0.74, p = 0.46). We found a significant main effect of compression (b = 0.36, SE = 0.05, t(12260) = 7.36, p < .001), indicating that compression on speech streams led to a decreased response speed. However, this main effect was qualified by a significant interaction with attention (b = 0.69, SE = 0.1, t(12260) = 7.11, p < .001). A closer examination of the interaction via post hoc tests showed that the decrease in response speed on compressed speech was driven by ignoring (b = -0.7, SE = 0.1, t(12260) = -7.35, p < .001) not attending (b = -0.012, SE = 0.017, t(12260) = -0.7, p = 1). That is participants showed slower responses to false alarms in the ignore compressed stream.

Comparing the hit rate between attend uncompressed and ignore compressed vs. attend uncompressed and ignore uncompressed, we found no significant difference (b = -0.09, SE = 0.07, OR = 0.92, 95% CI [0.8, 1.01], p = .22) as well as for attend compressed and ignore uncompressed vs. attend uncompressed and ignore uncompressed (b = -0.05, SE = 0.07, OR = 0.95, 95% CI [0.82, 1.1], p = .48). Participants detected significant less repeats correctly when both streams were compressed vs. both streams uncompressed (b = -0.25, SE = 0.07, OR = 0.78, 95% CI [0.67, 0.89], p < .001; Figure 16 A right).

We found a similar pattern comparing the response speed. No significant differences between attend uncompressed and ignore compressed vs. attend uncompressed and ignore uncompressed (b = -0.012, SE = 0.01, t(12252) = -1.2, p = .24) and attend compressed and ignore uncompressed vs. attend uncompressed and ignore uncompressed were observed (b = -0.01, SE = 0.01, t(12252) = -1.01, p = .29). Participants were significant slower detecting repeats in the attended stream when both streams were compressed vs both streams uncompressed indicated by decreased response speed (b = -0.03, SE = 0.01, t(12252) = -2.7, p = .007; Figure 16 B right).

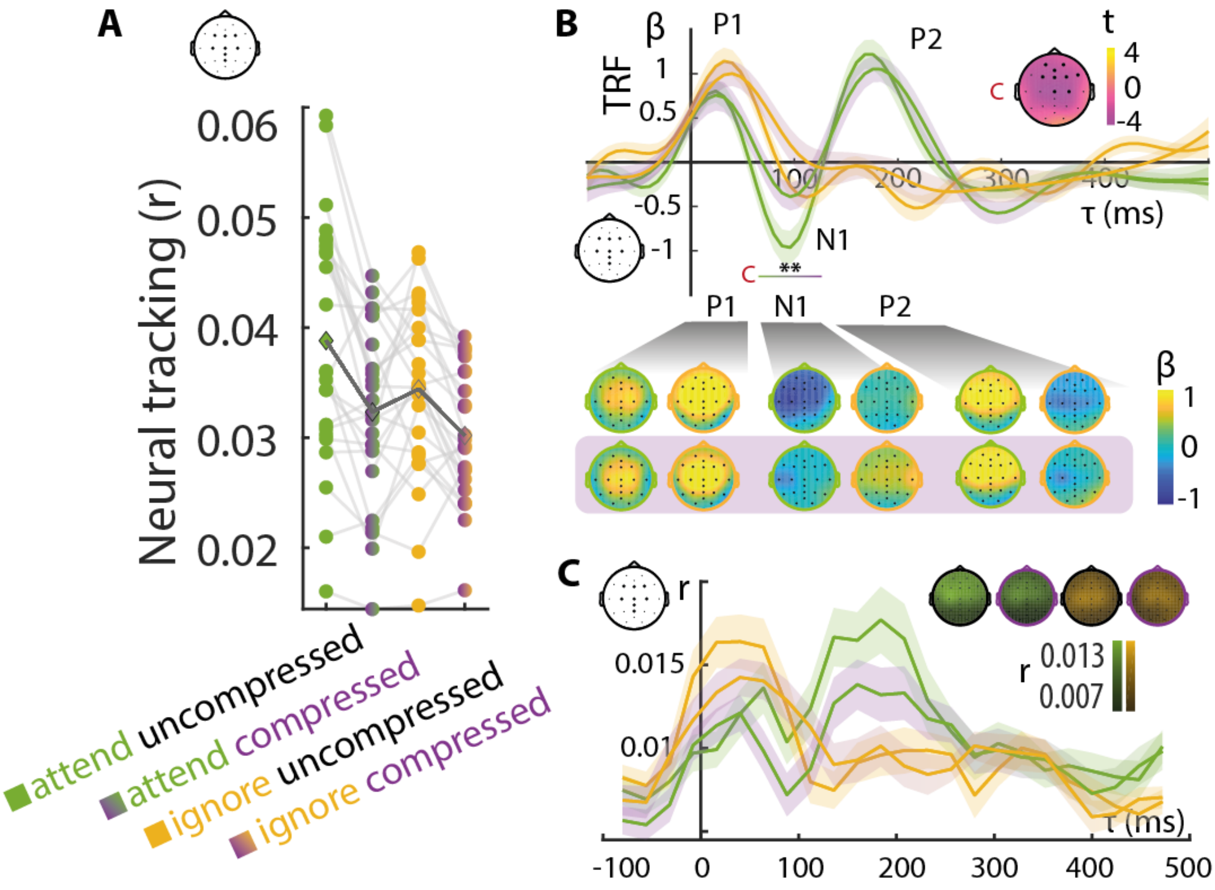


Figure 17: NH:Neural response

Neural tracking and TRF. A Neural tracking (r) refers to the encoding accuracy (0-500 ms) based on estimated TRFs and envelopes. Spaghetti plot shows single-subject data averaged across channels of interest. Connection lines between dots indicate the same subject. Green dots indicate attend tracking, while orange dots indicate ignored tracking. Purple color indicates compression. **B** TRF β-weights are averaged across subjects and channels of interest. Shaded areas depict the standard error for each time lag across subjects. Topographic maps depict β-weights for time windows of the P1, N1 and P2/N2 components. Topographic map (C) indicates and solide line (C) indicates significant cluster between attended uncompressed and attended compressed TRF. **C** Unfolding neural tracking across time lags (-100-500 ms). Solid lines shows the averaged neural tracking (r) across subjects and channels of interest (topographic map). Shaded areas show the standard error for each time lag across subjects.Topographic maps depict average neural tracking (r; 0-500 ms).

4.4.2 Compression decreases neural speech tracking

The strength of a speech stream's representation in the EEG is reflected by neural tracking (see methods for details). Analysis of the neural tracking (0-500 ms) revealed significant main effects of attention, which means that attended speech is stronger tracked than ignored speech (b = 0.08, SE = 0.02, t(9185) = 3.9, $p < .001$) and compression, which means that uncompressed speech is stronger tracked than compressed speech (b = 0.13, SE = 0.02, t(9185) = 6.4, $p < .001$; Figure 17 A). Attention and compression had no significant interaction (b = -0.06, SE = 0.04, t(9185) = -1.36, p =.175).

To assess the statistical significance of differences in TRFs between conditions, we used a cluster permutation test (Figure 17 B). We found a significant negative cluster (24–104 ms; cluster p-value < .001) which indicates a larger N1 amplitude for the attended signal, and a significant positive cluster (136-248 ms;

cluster p-value < .001) which indicates a larger P2 amplitude for the attended signal between attended uncompressed and ignored uncompressed TRFs. When attend compressed and ignore compressed were compared, there was a significant negative (40-96 ms; cluster p-value =.005) and a significant positive (144-248 ms; cluster p-value = <.001) cluster. We found a significant negative cluster comparing attended uncompressed and attended compressed TRFs (56-112 ms; cluster p-value = .003) which indicates a larger N1 amplitude for the attended uncompressed signal. No significant clusters between ignore uncompressed and ignore compressed were observed.

To assess the statistical significance of differences in time-shifted neural tracking between conditions, we used a cluster permutation test (Figure 17 C). We found a significant positive cluster between attend uncompressed and ignore uncompressed (136-208 ms; cluster p-value = .001). Comparing attend compressed and ignore compressed revealed a significant negative (64-88 ms; cluster p-value = .04) and a significant positive cluster (136-160 ms; cluster p-value = .04). We observed no significant clusters between attend uncompressed and attend compressed, as well as between ignore unprocessed and ignore compressed.

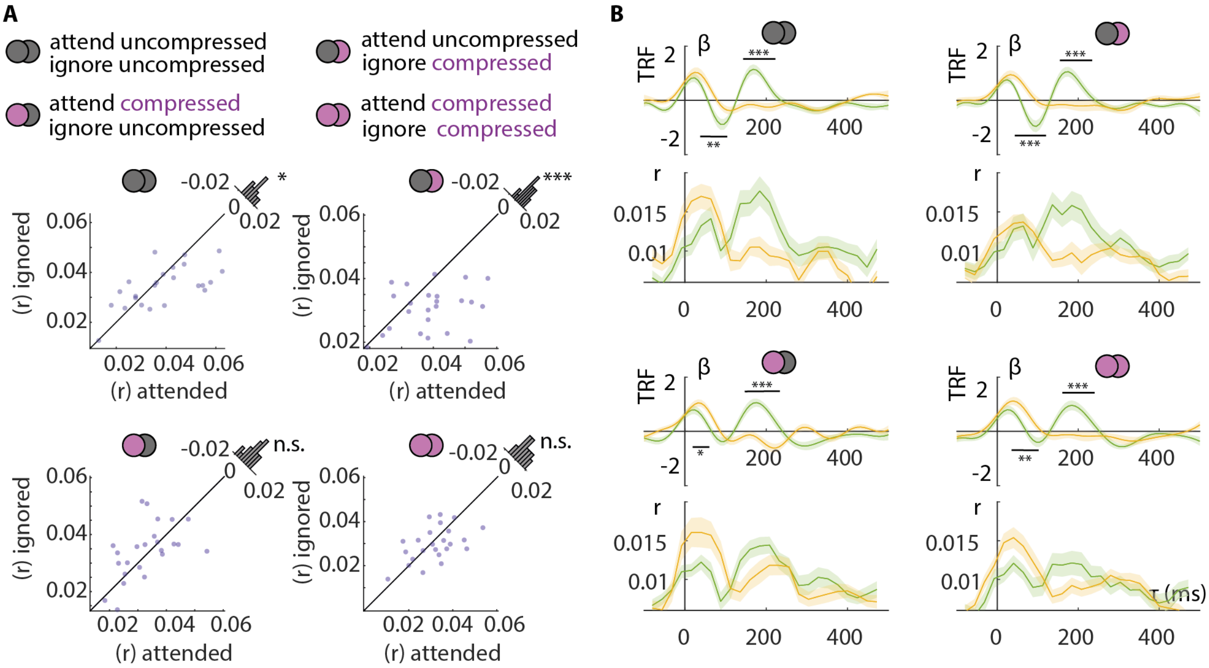


Figure 18: NH:Neural response pairs

Neural tracking and TRF within attentional pairs. Overlapping double circle symbols indicate attentional pairs. A circle in the foreground indicates attending while a circle in the background indicates ignoring. Purple colour indicates compression. **A** Neural tracking (r) refers to the encoding accuracy (0-500 ms) based on estimated TRFs and envelopes. 45° plot shows single-subject data averaged across channels of interest for the four attentional pairs. **B** TRF β-weights and time-shifted encoding accuracy are averaged across subjects and channels of interest for the four attentional pairs. Shaded areas depict the standard error for each time lag across subjects. Black bars indicate significant differences between attend and ignore TRF.

4.4.3 Increased neural separation due to decreased neural tracking of ignored stream

Analysis of the neural tracking (0-500 ms) within attentional pairs revealed (Figure 18 A) a significant effect of attention when both streams were uncompressed (b = 0.11, SE = 0.046, t(9208) = 2.58, p = 0.01). When only the ignored stream is compressed, we see significantly more neural tracking of the attended stream (b = 0.2, SE = 0.04, t(9208) = 4.9, p < .001). There was no significant difference between attend compressed and ignore uncompressed (b = -0.03, SE = 0.04, t(9208) = -0.83, p = .406) and when both streams were compressed (b = 0.05, SE = 0.04, t(9208) = 1.1, p = .263).

To assess the statistical significance of differences in TRFs within pairs, we used a cluster permutation test (Figure 18 B). We found one negative and one positive cluster (PC, NC) for each attentional pair: attend uncompressed & ignore uncompressed (NC: 40-104 ms, cluster p-value = .002; PC: 136-232 ms, cluster p-value < .001), attend uncompressed & ignore compressed (NC: 40-112 ms, cluster p-value = < .001; PC: 144-248 ms, cluster p-value < .001), attend compressed & ignore uncompressed (NC: 32-88 ms, cluster p-value = .03; PC: 128-240 ms, cluster p-value < .001) and attend compressed & ignore compressed (NC: 40-104 ms, cluster p-value = .002; PC: 152-248 ms, cluster p-value < .001).

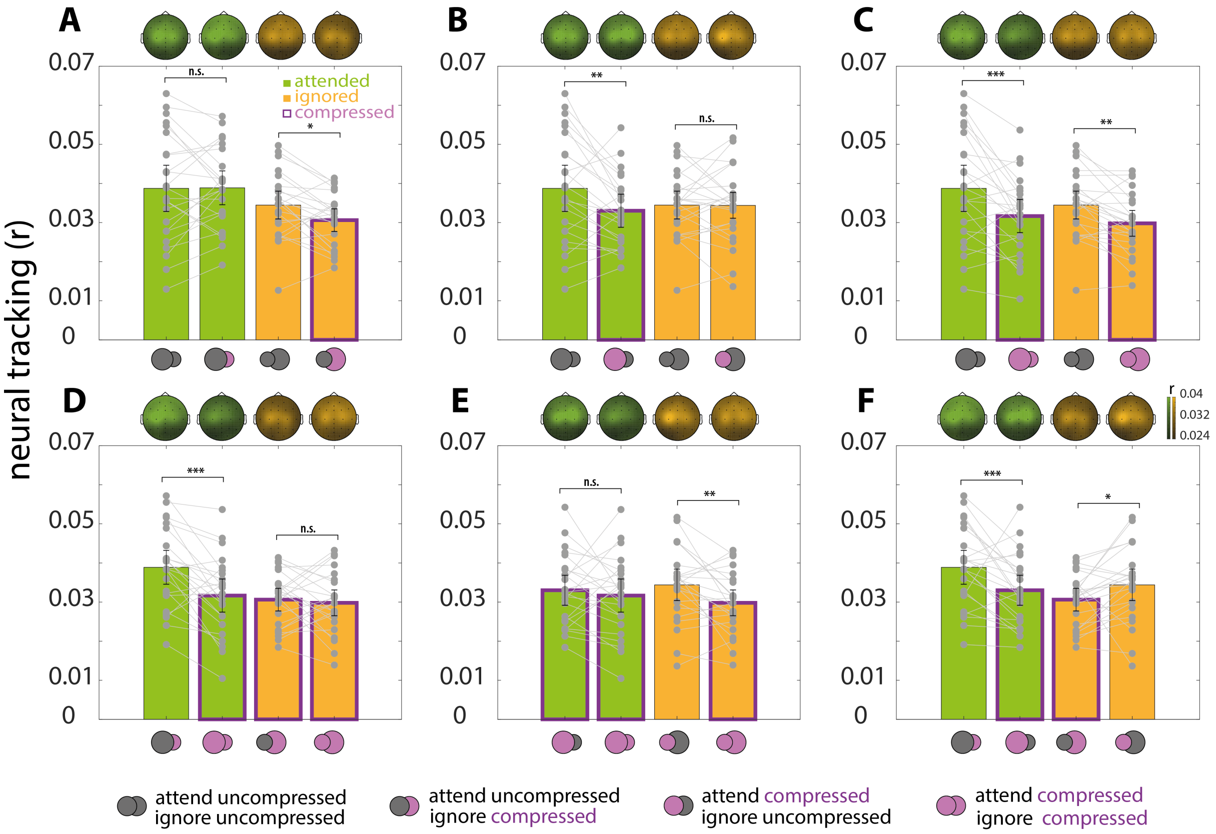


Figure 19: NH: Comparison attentional pairs

Comparison between attended and ignored neural tracking arising from different simultaneously presented attentional pairs. Neural tracking (r) refers to the encoding accuracy (0-500 ms) based on estimated TRFs and envelopes. Overlapping double circle symbols indicate attentional pairs. The purple color indicates compression. The bar plot shows the participant's group mean (green = attend, orange = ignore), and single-subject data are depicted as gray circles. Connection lines indicate the same subject. Error bars show a 95% CI. Topographic maps (green = attend, orange = ignore) depict average neural tracking (0-500 ms). **A** shows the comparison between attended and ignored streams from attend and ignore uncompressed pairs vs. attend uncompressed and ignore compressed pairs, **B** shows attend and ignore uncompressed pairs vs. attend compressed and ignore uncompressed pairs, **C** shows attend and ignore uncompressed pairs vs. attend and ignore compressed pairs, **D** shows attend uncompressed and ignore compressed pairs vs. attend and ignore compressed pairs, **E** shows attend compressed and ignore uncompressed pairs vs. attend and ignore compressed pairs, and **F** shows attend uncompressed and ignore compressed pairs vs. attend compressed and ignore uncompressed pairs.

Importantly, based on our hypothesis that compression on the ignored stream increases the neural separation between the attended and ignored streams, we have taken a closer look on the attentional differences between attended uncompressed & ignore compressed vs. attended uncompressed & ignore compressed (Figure 19 A). We discovered no statistically significant differences between the attended uncompressed streams from both pairs (b = -0.003, SE = 0.04, t(9208) = -0.08, p = 0.9). We found a significant difference between the ignored uncompressed and ignored compressed stream (b = 0.09, SE = 0.04, t(9208) = 2.3, p = 0.02) which indicates less neural tracking for the compressed ignored stream.

The results of the other comparisons (Figure 19 B-F) between attend (att) and ignore (ign) are shown in the following table:

Variable	Estimate	SE	tStat	DF	pValue	95 % CI
B: att	-0.141	0.041	-3.451	9208	0.001	[-0.221, -0.061]
B: ign	-0.002	0.041	-0.046	9208	0.963	[-0.082, 0.078]
C: att	-0.175	0.041	-4.265	9208	<.001	[-0.255, -0.094]
C: ign	-0.115	0.041	-2.809	9208	0.005	[-0.195, -0.035]
D: att	0.178	0.041	4.343	9208	<.001	[0.097, 0.258]
D: ign	0.02	0.041	0.48	9208	0.625	[-0.06, 0.1]
E: att	0.033	0.041	0.814	9208	0.416	[-0.047, 0.114]
E: ign	0.113	0.041	2.763	9208	0.006	[0.033, 0.194]
F: att	0.144	0.041	3.529	9208	<.001	[0.06, 0.225]
F: ign	-0.09	0.041	-2.275	9208	0.023	[-0.173, -0.013]

Table 1: Mixed model coefficients for each comparison.

We also investigated the comparison between attend and ignore for the attentional pairs in TRFs and time-shifted neural tracking (Figure 18 B), just like for neural tracking (Figure 19 A-F). Cluster permutation test revealed no significant differences for the comparisons.

4.5 Study 4: Auditory peripheral modeling of compressed and uncompressed speech

Before we started measuring people with presbycusis, we modelled the auditory periphery (see 2.6 Peripheral auditory modeling and Verhulst et al., 2018). We did that because the loss of outer hair cells changes the response behaviour to sounds and to account for the potential confound that the cortical results are driven by the human auditory periphery. While the healthy human cochlear is highly sensitive and non-linear (compressive), the damaged cochlear has a poorer frequency selectivity and is overall more linear in their response (Oxenham & Bacon, 2003).

We employed a computational model of the human auditory periphery developed by Verhulst and colleagues (2018) to simulate model outputs to the speech signals used in our experiments presented at 65 dB SPL. We randomly selected 100 speech snippets from our uncompressed stimulus material and applied the same processing pipeline (including compression and loudness matching) as used in our main experiment to create a set of 100 compressed speech snippets and a corresponding set of 100 uncompressed snippets. We modeled the firing rate of the auditory nerve (AN) and the envelope following response (EFR) for both normal hearing and hearing-impaired participants, simulating a typical mild-to-moderate presbycusis (hearing loss due to aging) starting at 1 kHz and sloping to 35 dB HL at 8 kHz. As the AN response varies with frequency, we focused on four center frequencies (500, 1000, 2000, and 4000 Hz) that are particularly relevant to speech in audiology (Sweetow & Silverman, 1994). The EFR, which reflects the neural processing of the temporal envelope, was modeled without frequency dependence.

To analyze the output of the auditory nerve (AN) in greater detail, we employed a mixed model. The dependent variable was the log-transformed spike rate of the modeled AN. We used the same speech snippet to generate both uncompressed and compressed outputs, for both normal hearing (NH) and hearing-impaired (HI) conditions, resulting in four different AN outputs for each speech snippet. To account for the quasi-repeated measures nature of the data, we included speech snippet as a random effect in the mixed model. In addition to hearing impairment (NH, HI) and signal manipulation (uncompressed, compressed), we also included frequency (500, 1000, 2000, and 4000 Hz) as a factor in the model.

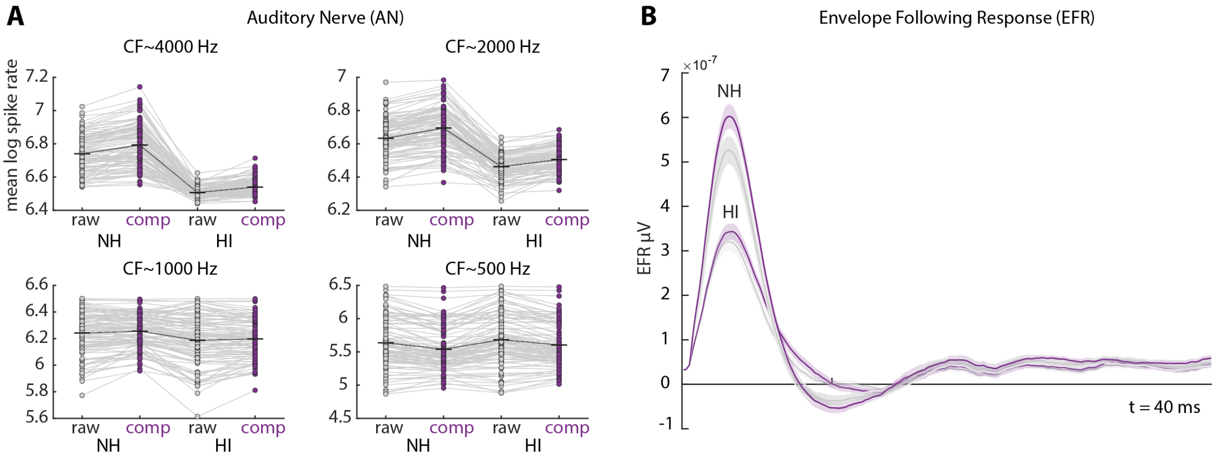


Figure 20: Verhulst modeling: AN and EFR

Simulated model output: AN and EFR. Panel **A** displays AN outputs for four center frequencies (500, 1000, 2000, and 4000 Hz), separated by normal hearing (NH) and hearing-impaired (HI) conditions and uncompressed (raw) and compressed (comp) speech snippets. Dots indicate different speech snippets. Connection lines indicate same snippet. Panel **B** shows the simulated EFR for NH and HI conditions,separate for compressed and uncompressed speech snippets. Shaded area shows SEM of different snippets, while solide line shows the mean across snippets. The color purple indicates compression.

We used a mixed model to analyze the AN output (Figure 20 A), with hearing status (NH vs. HI), signal manipulation (uncompressed vs. compressed), and frequency (500, 1000, 2000, and 4000 Hz) as fixed effects, and speech snippet as a random effect. Our analysis revealed a significant main effect of hearing status (b = 0.053, SE = 0.02, t = 2.4, p = 0.016), indicating higher log-transformed spike rates in the NH group compared to the HI group. We also observed a significant main effect of frequency (ref:1000 Hz; b = -1.3 to 0.9, SE = 0.03, t = -42 to 29, p < .001 for all), indicating higher log-transformed spike rates for higher frequencies. There was no significant main effect on signal manipulation (b = 0.004, SE = 0.02, t = .2, p = 0.9) In addition, there was a significant interaction between frequency and signal manipulation (overall, p = 0.003). No other interactions were significant (all p > 0.05). Overall, our results indicate that both hearing status and frequency have a significant impact on AN output.

Upon visual inspection the EFR (Figure 20 B), if at all, the compressed speech snippets appeared to lead to a higher amplitude in both simulations for normal hearing and hearing impaired.

The model outputs confirmed our expectation that the NH group has higher log-transformed spike rates than the HI group. The model simulates sensorineural hearing loss at the level of the transmission-line cochlear model and cochlear gain loss by reducing the gain of OHCs, which affects the model's tuning and sensitivity, leading to decreased AN model output for HI. The model also showed that the firing rate of the AN is higher for higher frequencies, in line with the literature (Greenwood, 1990). Although one would expect a stronger decrease in firing rate with increasing frequency for the HI simulation, the mixed model did not reveal a significant interaction between frequency and hearing status, but it was close to being significant (p = 0.051). Importantly, there was no significant effect of signal manipulation on the firing rate, and if anything, the amplitude was increased for loudness-matched compressed signals.

Interestingly, our previous neural tracking results in NH participants showed that compression reduced the neural tracking of compressed speech, which is the opposite of what was found in the simulated EFR. In conclusion, our simulation results suggest that the effects observed in our neural speech tracking study are not confounded due to changes in the auditory periphery. To further investigate the impact of hearing loss on neural speech tracking of compressed speech, we conducted a similar experiment with HI participants.

4.6 Study 5: The interplay between dynamic range compression and selective attention in hearing impaired listeners

Dynamic range compression is commonly used in nearly all modern hearing aids to compensate for loudness recruitment and is generally considered to have a positive effect on users' hearing. However, the results in the literature are mixed with some studies finding benefit for compression and other studies finding no benefit or even a detriment to intelligibility or speech quality (Braida et al., 1979; Dillon, 1996; Souza, 2002). We here used a relative strong compression ratio of 1:8 so we would not expect positive effects of compression but rather decline in behavioural performance and neural response similar to the normal hearing. There is also some evidence for the hypothesis that hearing impaired people use cortical neural tracking to compensate for their peripheral heraring loss (Schmitt et al., 2022). Hence we hypothesize that our choosen compression parameters lead an even larger neural seapartion between attended and compressed ignored speech in contrast to normal hearing participants.

In this study, 7 participants with presbycusis had their electroencephalograms (EEGs) recorded while they listened to two different narrated audio streams. The level of compression applied to each stream was randomly assigned to each participant. Participants had to switch between the two streams and identify any repeated segments in the attended stream while ignoring the unattended stream. Although the double iterative bootstrapping analysis performed in this study allowed us to investigate differences between conditions in a group of seven hearing-impaired participants, it is important to note that the sample size is small and the results should be interpreted with caution. Additionally, it should be noted that bootstrapping here is always a pairwise comparison and may not be sensitive to interactions between conditions or other factors.

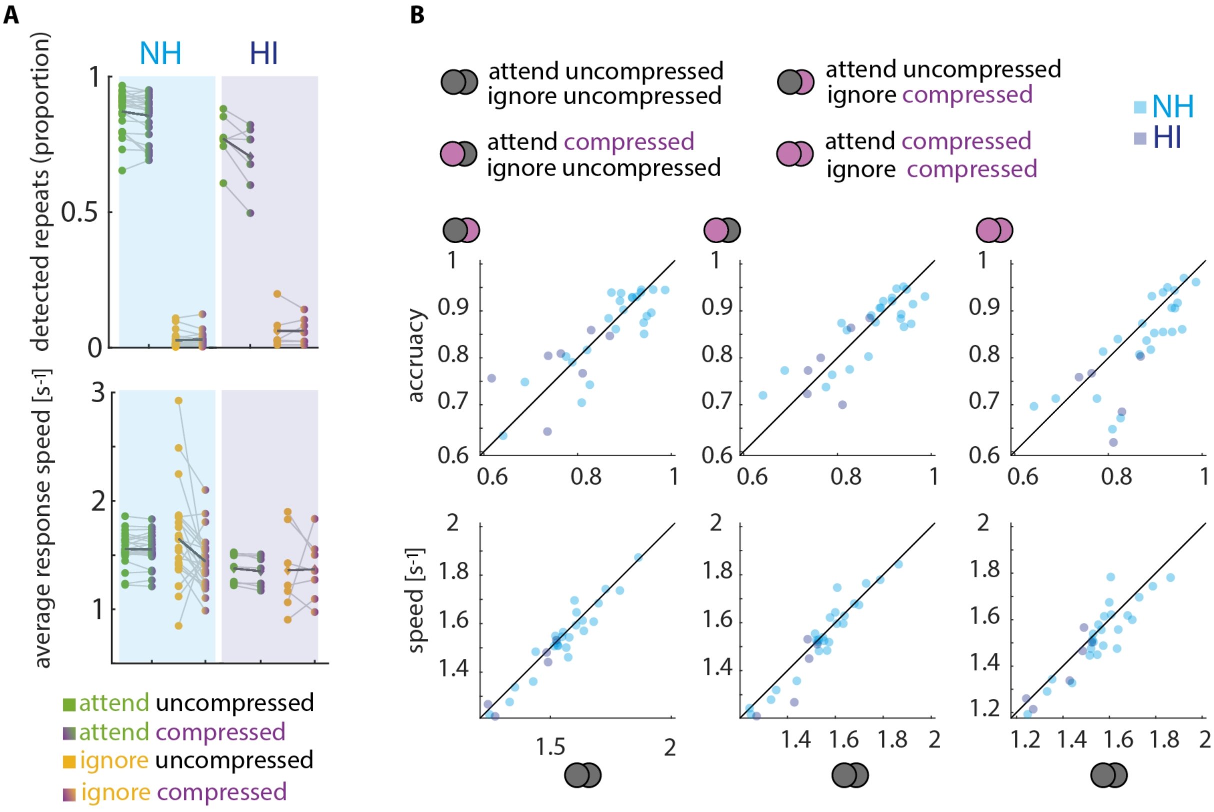


Figure 21: HI: Behavioural results

Behavioral results for participants with hearing impairment (HI, N=7) and normal hearing controls (NH, N =24). In panel **A** the upper section displays the proportion of detected repeats per condition, with individual mean proportions shown as dots, and light gray lines indicating the same subject. The mean of the group is represented by the bold line. The lower section shows response speed per condition, with individual mean response speeds displayed as dots and light gray lines indicating the same subject. The mean of the group is represented by the bold line. Panel **B** shows the upper section with the proportion of correctly detected repeats in the attend stream for attentional pairs (attend/ignore) with the four experimental compression combinations. The attentional pair (attend-uncompressed:ignore-uncompressed) serves as a baseline and is compared against the remaining three pairs, with individual subject data depicted as dots. The lower section displays response speed of correctly detected repeats in the attend stream for attentional pairs (attend/ignore) with the four experimental compression combinations, with the same baseline comparison and individual subject data depicted as dots.

4.6.1 Slower responses to attended compressed speech

The study found that participants with hearing impairment (HI) were able to detect repeated sounds in both the uncompressed and compressed attended streams, with mean accuracies of 0.77 (95% double bootstrapped [CI]: [0.63, 0.86]) and 0.7 (95% double bootstrapped CI: [0.56, 0.81]), respectively (Figure 21 A). The difference in accuracy between the uncompressed and compressed attended streams was not significant (mean difference: 0.07, 95 % double bootstrapped CI:[-0.015, 0.128]). HI participants also made very few false alarms in detecting repeated sounds in the ignored uncompressed and compressed streams, with mean accuracies of 0.14 (95% double bootstrapped CI: [0.02, 0.15]) and 0.06 (95% double bootstrapped CI: [0.02, 0.12]), respectively.The difference in false alarms between the compressed and uncompressed ignored streams was not significant (mean difference: -0.001, 95% double bootstrapped CI: [-0.04, 0.04]. These results suggest that there is no significant difference in performance within the HI group between the compressed and uncompressed attended or ignored streams.

Participants with hearing impairment (HI) responded faster to repeated sounds in the uncompressed attended stream compared to the compressed attended stream, with mean response speed of 1.37 s^{-1} (95% double bootstrapped confidence interval [CI]: [1.26, 1.5]) and 1.35 s^{-1} (95% double bootstrapped CI: [1.23, 1.48]), respectively (Figure 21 B). The difference in response speed between the uncompressed and compressed attended streams was significant (mean difference: 0.027 s^{-1}, 95% double bootstrapped CI: [0.045, 0.05]). However, there was no significant difference in response speed between the uncompressed and compressed ignored streams, with mean response times of 1.37 s^{-1} (95% double bootstrapped CI: [0.97, 1.86]) and 1.37 s^{-1} (95% double bootstrapped CI: [1.05, 1.79]), respectively (mean difference: -0.01 s^{-1} , 95% double bootstrapped CI: [-0.58, 0.35]). These results suggest that there is a significant difference in performance within the HI group between the uncompressed and compressed attended streams, with faster response times for the uncompressed stream.

The results of the double iterative bootstrapping analysis for attentional pairs indicated that there were no significant differences between accuracy and speed in hearing-impaired participants.

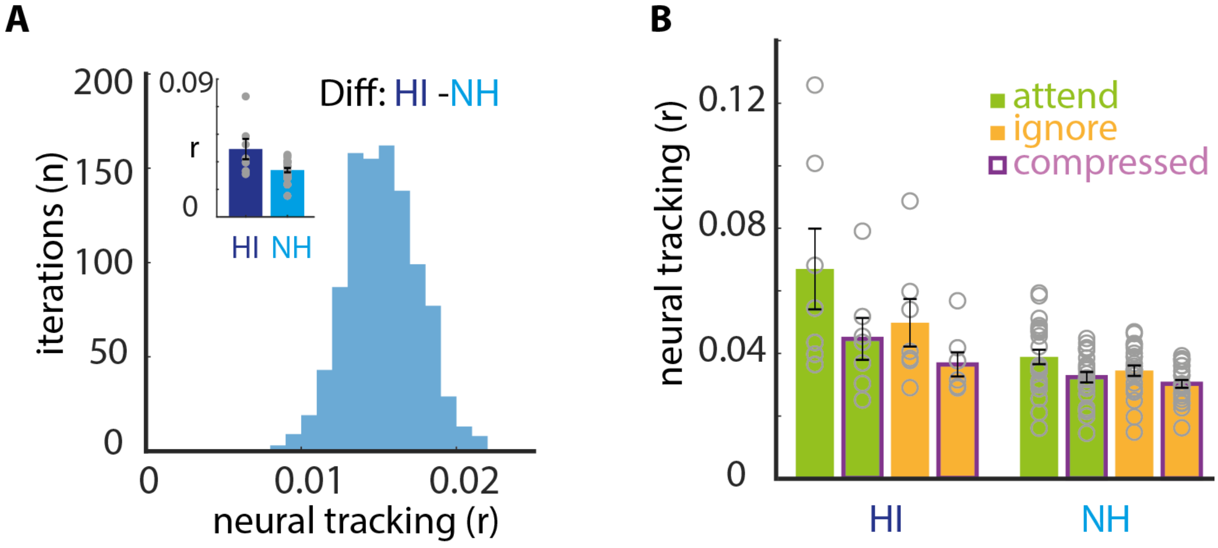


Figure 22: HI: Neural response

Neural tracking for participants with hearing impairment (HI, N=7) and normal hearing controls (NH, N =24). In Panel **A** of the figure, the differences in neural tracking between the hearing impairment (HI) group (consisting of 7 participants) and normal hearing (NH) control group (consisting of 24 participants) were obtained by selecting 7 participants from the NH group and comparing their neural tracking to that of the HI group. This process was repeated 1000 times to obtain an average difference between the two groups. The inset shows the average neural tracking across all conditions for both groups. Group averages are represented by bars and individual participant averages by dots. Panel **B** shows the neural tracking for each condition for both groups, with group averages represented by bars and individual participant averages by circles. Error bars show the standard error (SEM).

4.6.2 Hearing impaired show stronger neural tracking than normal hearing

We compared neural tracking between a group of 24 normal hearing control participants and a group of 7 hearing-impaired participants (Figure 22 A). Neural tracking was averaged across all conditions and the differences in neural tracking between the groups were obtained by randomly selecting 7 participants from the normal hearing group and comparing their neural tracking differences to that of the hearing-impaired group. This process was repeated a thousand times to obtain a distribution of the iterative differences between the two groups. The results suggest that the overall neural tracking was larger for the hearing-impaired group compared to the normal hearing control group. Specifically, all iterative differences were larger than zero.

4.6.3 Compression decreases neural tracking in hearing impaired

Despite the small sample size of hearing-impaired participants, the neural tracking results are consistent (Figure 22 B). All participants showed increased neural tracking when they attended to signals compared to when they ignored them (mean: 0.013, 95% double bootstrapped CI: [0.001, 0.0335]). They also demonstrated significantly less neural tracking of compressed speech compared to uncompressed speech (mean: 0.018, 95% double bootstrapped CI: [0.0075, 0.0317]). Furthermore, the difference in neural tracking between attending to uncompressed versus compressed speech was significant (mean: 0.022, 95% double bootstrapped CI: [0.0029, 0.0439]), as was the difference in neural tracking between ignoring uncompressed versus compressed speech (mean: 0.014, 95% double bootstrapped CI: [0.0014, 0.0256]).

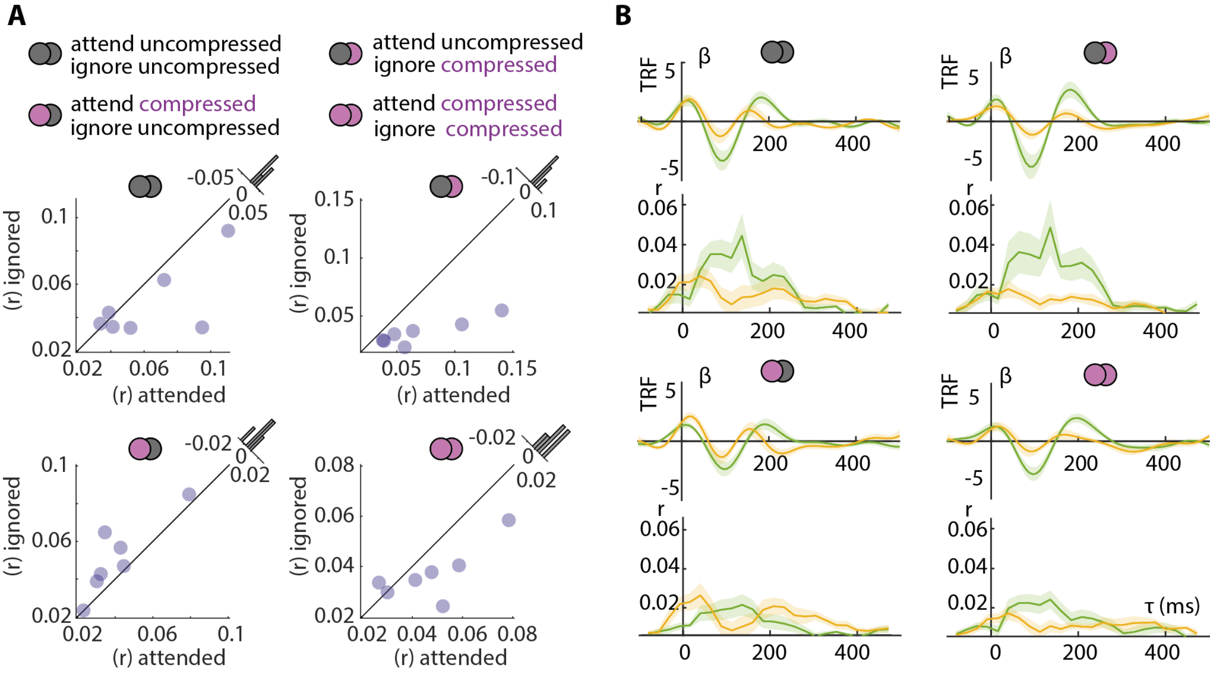


Figure 23: HI: Neural response pairs

Neural tracking and TRF to attentional pairs for hearing impaired participants. Overlapping double circle symbols indicate attentional pairs. A circle in the foreground indicates attending while a circle in the background indicates ignoring. Purple colour indicates compression. **A** Neural tracking (r) refers to the encoding accuracy based on estimated TRFs and envelopes. 45° plot shows single-subject data averaged across channels of interest for the four attentional pairs contrasted between attended and ignored. **B** TRF β-weights and time-shifted encoding accuracy are averaged across subjects and channels of interest for the four attentional pairs (green = attended, orange = ignored). Shaded areas depict the standard error for each time lag across subjects.

4.6.4 Largest neural separation in the attend uncompressed - ignore compressed attentional pair

We conducted an analysis of neural tracking (0-500 ms) within attentional pairs (attend vs. ignore; Figure 23 A) and found no significant difference when both streams were uncompressed (mean: 0.015, 95% double bootstrapped CI: [-0.0037, 0.0475]). However, we did observe a significant difference in the attentional pair when only the ignored stream was compressed (mean: 0.035, 95% double bootstrapped CI: [0.0095, 0.0750]). Interestingly, when the attended stream was compressed and the ignored stream was uncompressed, we found significantly less neural tracking of the attended stream (mean: -0.0102, 95% double bootstrapped CI: [-0.0256, -0.0011]). When both streams were compressed, there was no significant difference between the two streams (mean: 0.0109, 95% double bootstrapped CI: [-0.0024, 0.0243]).

Upon visual inspection of the temporal response functions (TRFs) and time-shifted neural tracking, we found that they were consistent with the neural tracking results (Figure 23 B). Overall, in attentional pairs, the attended TRFs showed the typical P1-N2-P2 succession. Additionally, the P1 component of both the attended and ignored TRFs had similar amplitudes across all attentional pairs. However, the difference between the attended and ignored streams varied between attentional pairs for the N1 time window (approximately 80-120 ms). The largest separation occurred in the attend uncompressed and ignore compressed pair, followed by the attend and ignore uncompressed and attend and ignore compressed pairs. Interestingly, there appeared to be no separation in the attend compressed and ignore uncompressed pair. These findings were consistent with the time-shifted neural tracking results, which indicated similar neural separation relations in attentional pairs.

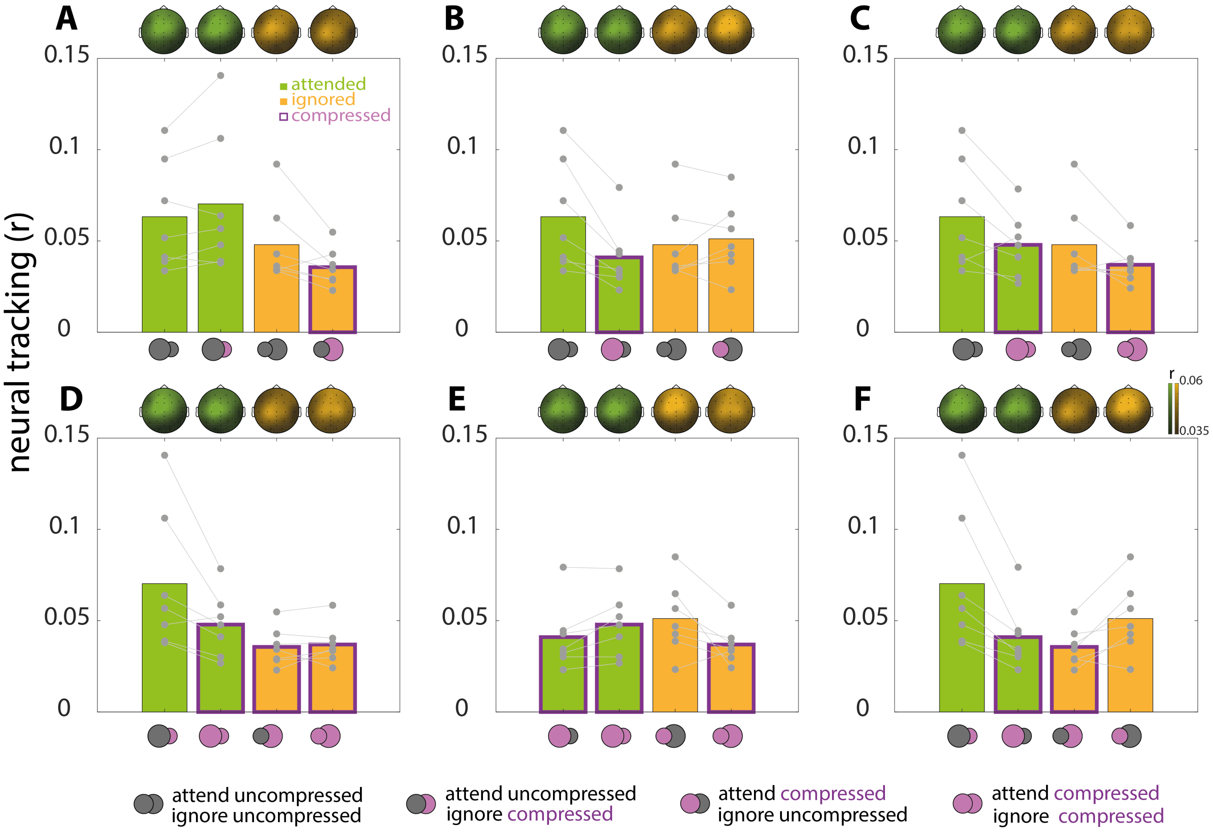


Figure 24: HI: Comparison attentional pairs

Comparison between attended and ignored neural tracking arising from different simultaneously presented attentional pairs. Neural tracking (r) refers to the encoding accuracy 0-500 ms based on estimated TRFs and envelopes. Overlapping double circle symbols indicate attentional pairs. The purple color indicates compression. The bar plot shows the participant's group mean (green = attend, orange = ignore), and single-subject data are depicted as gray circles. Connection lines indicate the same subject. Topographic maps (green = attend, orange = ignore) depict average neural tracking (0-500 ms). **A** shows the comparison between attended and ignored streams from attend and ignore uncompressed pairs vs. attend uncompressed and ignore compressed pairs, **B** shows attend and ignore uncompressed pairs vs. attend compressed and ignore uncompressed pairs, **C** shows attend and ignore uncompressed pairs vs. attend and ignore compressed pairs, **D** shows attend uncompressed and ignore compressed pairs vs. attend and ignore compressed pairs, **E** shows attend compressed and ignore uncompressed pairs vs. attend and ignore compressed pairs, and **F** shows attend uncompressed and ignore compressed pairs vs. attend compressed and ignore uncompressed pairs.

Figure 24 shows the comparisons between the attend and ignore neutral tracking conditions for attentional pairs in hearing-impaired participants, just like Figure 19 for normal hearing participants. The difference in Figure 24 A will be analyzed in more detail in the next paragraph and compared to the results for normal-hearing participants. No further statistical analyses were conducted for the other comparisons. However, based on visual observation, it can be suggested that the general pattern is similar to that of normal-hearing participants, namely that amplitude compression reduces neural tracking.

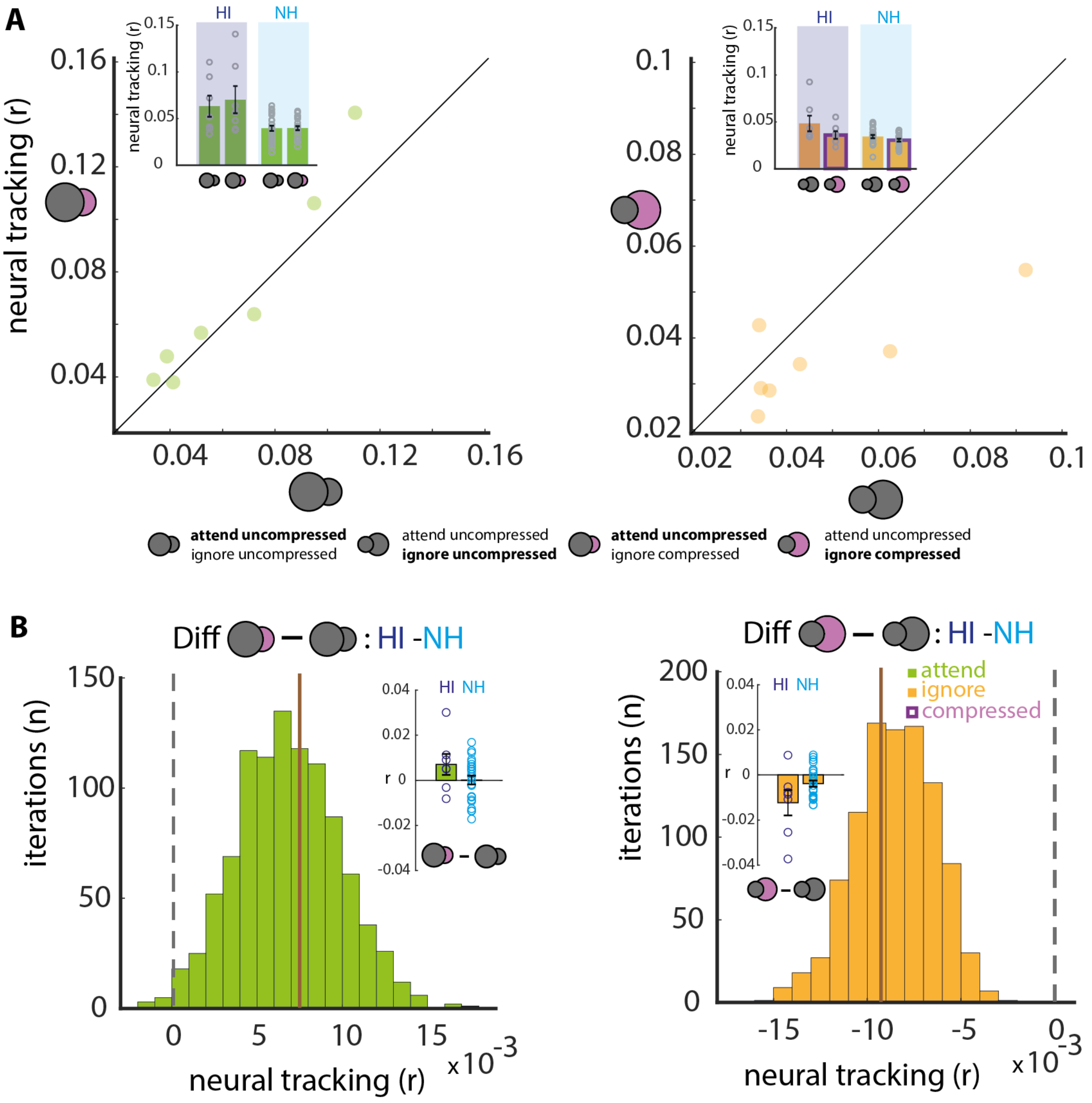


Figure 25: Neural response HI vs. NH

Comparing neural tracking of attended and ignored uncompressed versus attended uncompressed and ignored compressed between hearing impaired (HI) and normal hearing control (NH). Overlapping double circle symbols indicate attentional pairs. Purple colour indicates compression. **A** 45° plot shows single-subject data averaged across channels of interest for the attentional pairs contrasted between attended and ignored. Inset: Bar plot shows participant's mean and single-subject data are depicted as circle. Dark blue area indicates HI and light blue area NH participants. Error bars show the standard error (SEM). **B** Comparison of the neural tracking differences between the attended streams (left) and ignored streams (right) between hearing impaired (HI) and normal hearing (NH) participants for selected pairs. Distributions were obtained by randomly selecting seven participants from the NH group and comparing their neural tracking to that of the HI group. This process was repeated 1000 times to obtain an average difference between the two groups. The solid vertical lines (brown) indicate the average difference across iterations, while the dashed line indicates zero, which would indicate no difference between HI and NH. Inset shows differences in neural tracking between the attended and ignored streams for the attentional pair comparison, respectively. Bar plot shows participant's mean and single-subject data are depicted as circle. Dark blue circle indicates HI and light blue circle NH participants. Error bars show the standard error (SEM).

4.6.5 Hearing impaired show larger differences between attended and ignored streams compared to normal hearing

To investigate whether the neural separation is larger in the hearing impaired when the ignored stream is compressed, we compared the attentional pairs in which both streams are unprocessed to the pair in which only the ignored stream is compressed and the attended stream is uncompressed. We compared the attended (mean: -0.007, 95% double bootstrapped CI: [-0.0244, 0.0068]) and ignored (mean: 0.0123, 95% double bootstrapped CI: [-0.0043, 0.0322]) streams between the ”attend uncompressed and ignore uncompressed” pair and the ”attend uncompressed and ignore compressed” pair. The results showed no significant difference between the two pairs. However, visual inspection of the individual participants' data revealed that most (5 out of 7) participants showed increased neural tracking for attend , and most (6 out of 7) showed decreased neural tracking for ignore for the ”attend uncompressed ignore compressed” attentional pair (Figure 25 A) .

We compared the differences in neural tracking between the attended and ignored speech streams for two attentional pairs: ”attend uncompressed and ignore uncompressed” and ”attend uncompressed and ignore compressed”, between a group of 24 normal hearing control participants and a group of 7 hearing-impaired participants (Figure 25 B). To obtain the differences in neural tracking between the two groups, we randomly selected 7 participants from the normal hearing group and compared their neural tracking differences to that of the hearing-impaired group. This process was repeated a thousand times to obtain a distribution of the iterative differences between the two groups.

We found that the differences in attended streams between the groups were mostly larger for the hearing-impaired participants (mean: 0.0069, 95% CI: [0.067, 0.070]). However, the differences in ignored stream between the groups were consistently larger in the hearing-impaired listeners across all iterations (mean: -0.0086, 95% CI: [-0.0084, -0.0087]).

4.6.6 Control analysis: Front back location assignment does not confound neural and behavioural results

We considered the possibility that the front back location assignment could have an indirect effect on our behavioural and neural measures. Between trials (and for some sustained trials), participants had to switch their attention between the front and back loudspeakers. We randomized and balanced our conditions across the two locations of the streams. Nevertheless, we used the location as a factor in our statistical analysis to control for potential confounds.

In our study, we analyzed the effects of location (front-back), attention, and compression on both behavioural performance and response time in normal hearing participants. Our results showed a significant main effect of location on behavioural performance (b = 0.26, SE = 0.06, OR = 1.3, 95% CI [1.16, 1.45], p < .001), indicating that participants detected more repeats in the front loudspeaker.

Importantly, we found no significant interactions between location and attention (b = -0.01, SE = 0.11, OR = 0.99, 95% CI [0.79, 1.24], p = 0.9), between location and compression (b = -0.14, SE = 0.11, OR = 0.87, 95% CI [0.7, 1.01], p = 0.2), and between location, compression, and attention (b = -0.22, SE = 0.23, OR = 0.8, 95% CI [0.51, 1.26], p = 0.33).

Regarding response speed, we observed no significant main effect of location (b = 0.02, SE = 0.05, t = 0.44, p = 0.66) and no significant interactions between location and attention (b = -0.08, SE = 0.1, t = -0.8, p = 0.43), between location and compression (b = -0.01, SE = 0.1, t = -0.1, p = 0.92), and between location, compression, and attention (b = -0.02, SE = 0.19, t = -0.09, p = 0.93).

Our neural analysis showed no significant main effect of location (b = -0.02, SE = 0.02, t = -1.04, p = 0.3), and we found no significant interactions between location and attention (b = -0.03, SE = 0.04, t = -0.8, p = 0.43), between location and compression (b = -0.06, SE = 0.04, t = -1.34, p = 0.175), or between location, compression, and attention (b = -0.03, SE = 0.08, t = -0.312, p = 0.76). Therefore, our results suggest that the front-back location assignment did not confound the neural and behavioural measures in normal hearing participants.

Hearing impaired participants correctly detected repeats on average with a proportion of 0.76 (95% CI [0.64, 0.88]) in the front loudspeaker, while they correctly detected a proportion of 0.71 (95% CI [0.63, 0.79]) in the back loudspeaker. They responded with similar response speed to the front (mean: 1.37 s^{-1}, 95% CI [1.24, 1.49]) and back loudspeaker (mean: 1.36 s^{-1}, 95% CI [1.23, 1.5]).

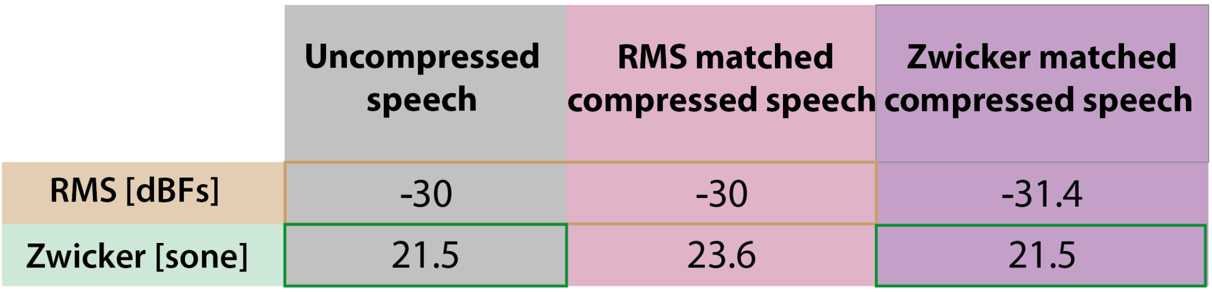
In our neural analysis, hearing impaired participants show a mean neural tracking of 0.06 (95% CI [0.03, 0.09]) for attended and 0.046 (95% CI [0.03, 0.06]) for ignored speech presented over the frontal loudspeaker. In contrast, they show a mean neural tracking of 0.052 (95% CI [0.03, 0.07]) for attended and 0.04 (95% CI [0.03, 0.05]) for ignored speech presented over the back loudspeaker.

4.7 Study 6: perceived loudness of compressed and uncompressed speech stimuli

Auditory stimulation level is a widely recognized confounding factor in psychoacoustics and auditory neurophysiology. As a result, nearly all studies related to hearing aim to control for this factor. To account for potential differences in perceived loudness between uncompressed and compressed speech, we applied a Matlab algorithm based on Zwicker's model (Zwicker & Scharf, 1965). Previous research has demonstrated that RMS-matching can increase the perceived loudness of compressed speech compared to uncompressed speech (Moore et al., 2003). While existing models can accurately predict the loudness of stationary signals, the accuracy of these models for time-varying signals depends on the specific stimulus; however, loudness perception of speech appears to be relatively robust (Rennies, Verhey, Appell, & Kollmeier, 2013). Nonetheless, to ensure that perceived loudness was comparable between the compressed and uncompressed speech stimuli used in our study, we conducted an online study to assess participants' actual loudness perception on our used stimuli material.

4.7.1 Participants

Ten participants (7 female and 3 male) between the ages of 22 and 32 with no reported hearing impairments participated in this online study. Eight of them were native German speakers. Participants used a PC with headphones, and the volume on each PC was adjusted to a level where soft speech signals could still be heard, while very loud speech signals were not uncomfortably loud. It should be noted that the volume range was set individually for each participant. In the experiment, 720 speech signals (40 different speech signals * 3 conditions * 6 volume levels) were randomly presented to the participants. After each presentation, participants rated the loudness of the speech signal on a scale of 1 to 9, with 1 indicating a soft signal and 9 indicating a loud signal. Participants were instructed to use the entire scale during the experiment.



	Uncompressed speech	RMS matched compressed speech	Zwicker matched compressed speech
RMS [dBFs]	-30	-30	-31.4
Zwicker [sone]	21.5	23.6	21.5

Figure 26: Example loudness matching

Example of RMS and Zwicker matching . The schematic shows an example of the RMS (dBFS) and Zwicker (sone) outputs of an uncompressed speech signal, which was used as a matching baseline and of the corresponding RMS and Zwicker matched signals. The brown rectangle indicates the pair of speech signals that were matched based on RMS, while the green rectangle indicates the pair that were matched based on Zwicker loudness, which were used in the experiment.

4.7.2 Methods

We randomly selected twenty 2.1-second speech snippets from our uncompressed stimulus material. These signal segments were then processed to generate two sets of stimuli: one set matched to six different RMS levels (-36, -33, -30, -27, -24, and -21 dBFS), and one set matched to six different loudness levels according to the Zwicker model (Figure 26; following the same processing pipeline as in our main experiment). These different levels were used to generate loudness functions using the method described by J. C. Stevens and Marks (1980),originally used to achieve cross-modality matching of loudness and brightness. For the online experiment, the speech signals were available in three different conditions: uncompressed, Zwicker-matched compressed, and RMS-matched compressed.

4.7.3 No significant differences in perceived loudness between uncompressed and Zwicker loudness matched compressed speech

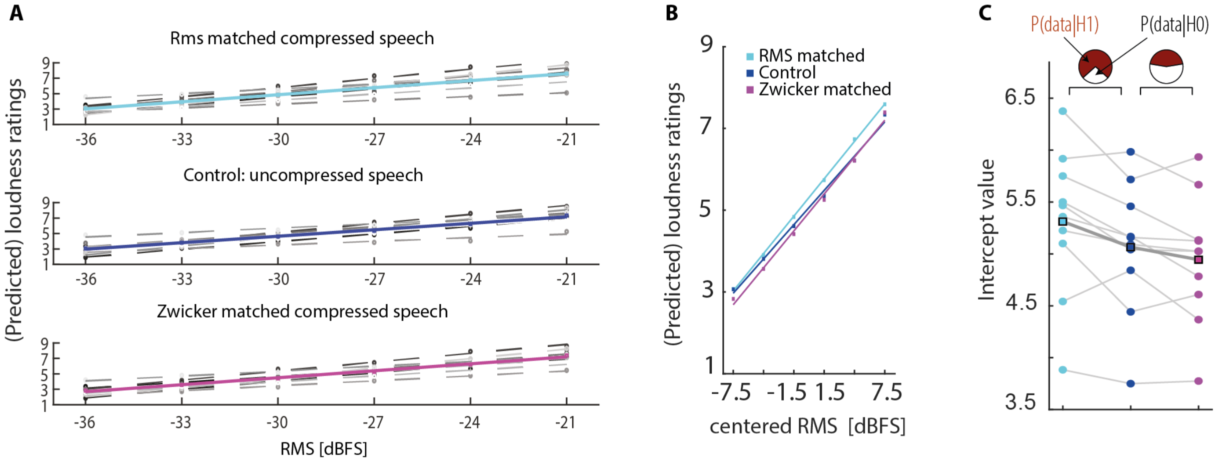


Figure 27: Loudness ratings

Results: Loudness ratings. Figure **A** shows the mean values of the different conditions per RMS value (circle) and the resulting linear regression for each subject (dashed line). Figure **B** shows this linear regressions of mean values across all subjects for each condition as a function of the centered RMS value in direct comparison. Figure **C** presents the calculated intercept values of each subject (circle) and the mean values across all subjects (square) for mean-adjusted RMS values. To visualize the Bayes factor, we have included probability pie charts in which the ratio of the likelihood of H1 (shown in red) and H0 (shown in white) for pairwise comparisons is displayed.

The results of the linear regressions showed a positive slope in all conditions, indicating that higher levels were perceived as louder. The study also found that each participant used almost the same range on the rating scale for all conditions, with some preferring the upper end of the scale and others the lower end (Figure 27 A).

Comparing the regression lines across all participants, it was found that the regression line for the loudness rating of the RMS-matched compressed speech signal was consistently above that of the uncompressed speech signal, indicating that the former was rated as louder. However, for the Zwicker-matched compressed speech signal, the regression line was below that of the uncompressed speech signal, although at higher levels the two were nearly superimposed, with the Zwicker-matched compressed speech signal slightly higher (Figure 27 B).

The results of the Bayesian t-test indicate that there is anecdotal evidence (BF10 = 2.709) supporting the hypothesis (H1) that the intercepts of the RMS matched compressed speech signals have different mean values compared to the control - uncompressed speech signals. This suggests that the RMS matched speech is perceived as louder. On the other hand, there is anecdotal evidence (BF10 = 0.814) supporting the null hypothesis (H0) that the intercepts of the Zwicker matched compressed speech signals do not have different mean values compared to the control - uncompressed speech signals. This indicates that there is similar loudness perception between these two conditions (Figure 27 C).

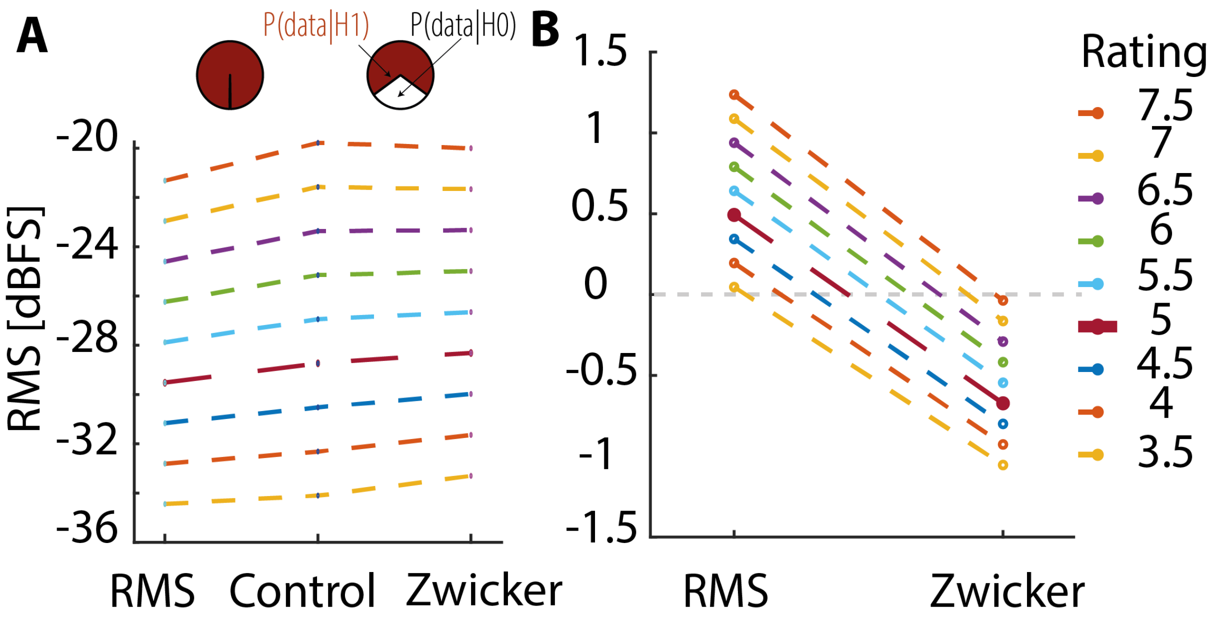


Figure 28: Loudness as function of conditions

Loudness as function of conditions. **A** For ratings 3.5 to 7.5, the RMS values were calculated via linear regression of each condition.To visualize the Bayes factor, we have included probability pie charts in which the ratio of the likelihood of H1 (shown in red) and H0 (shown in white) for pairwise comparisons is displayed. **B** The difference between the respective compressed conditions and the control - uncompressed condition. Dashed line represents control condition.

The Bayesian paired t-test produced strong evidence (BF10 = 236.19) indicating that there are differences in mean values between the dBFS values associated with the ratings of the RMS matched compressed speech signal and those of the control - uncompressed speech signal (H1). On the other hand, for the comparison between the Zwicker matched compressed speech signal and the control - uncompressed speech signal, the Bayes factor showed anecdotal evidence for H1 (BF10 = 2.31; Figure 28 A).

The comparison between the values of the compressed speech conditions and the control - uncompressed speech signal illustrates the extent to which the RMS value of the control - uncompressed speech signal needs to be increased or decreased to be perceived as equally loud, as depicted in Figure 5B. The differences for the RMS matched compressed speech signal range between 1.2 and 0 dB, whereas for the Zwicker matched compressed speech signal, the range is between 0 and -1.1 dB (Figure 28 B).

The aim of this study was to validate the loudness matching of uncompressed and compressed speech using a Zwicker-based algorithm, which also incorporated RMS matching, in a psychoacoustical experiment. The results suggest that there is a difference in loudness perception between the compressed speech signals and the control - uncompressed speech signal, with the RMS matched compressed speech signal is rather perceived as louder. Importantly, the results suggest no significant differences in loudness perception between uncompressed and Zwicker matched compressed speech. In addition, the difference of 1.1 dB found here can be considered very small, since the detection thresholds for level differences is around 1 dB (S. S. Stevens, Volkmann, & Newman, 1937).

4.8 Discussion

In the present study, we aimed to investigate the effects of amplitude compression and selective attention on neural separation and behavioral response in normal and hearing impaired participants using a psychophysically augmented continuous speech paradigm. Our hypothesis was that the interaction between attention and compression would affect both neural separation and behavioral response. Specifically, we expected that compression on ignored talkers would increase neural separation and behavioral response.

Normal hearing participants showed several key results. Firstly, when compression was applied to both attended and ignored streams, behavioral performance decreased, and response to repeats in the ignored stream was slower. Secondly, we observed a significant main effect of compression on decreased neural speech tracking. Finally, the most important finding was an increased neural separation in neural speech tracking between attentional pairs when when only the ignored stream was compressed compared to both streams compressed. This difference was driven by the decreased neural tracking of the compressed ignored stream.

For hearing impaired participants, we found slower responses to attended compressed speech and evidence that amplitude compression decreased neural tracking, which was a relatively robust finding given the small sample size. The most important result was that we observed the largest neural separation in the attentional pair where the attended stream was uncompressed and the ignored stream was compressed. This finding was not entirely unambiguous due to the small sample size. However, there was a tendency for hearing impaired participants to show both enhanced tracking of the uncompressed attended stream and weaker tracking of the compressed ignored stream.

When comparing the results of normal hearing and hearing impaired participants, we found that hearing impaired participants showed increased overall neural tracking compared to normal hearing participants. Additionally, hearing impaired participants showed a larger difference in neural tracking between the attended and ignored streams when both streams were compressed compared to when only the ignored stream was compressed compared to normal hearing participants.

4.8.1 Amplitude compression on both attend and ignore speech impairs performance

In the present study, normal hearing participants showed decreased behavioral accuracy and slower responses when the attended and ignored streams were simultaneously compressed compared to when both streams were uncompressed (Figure 16). The findings are consistent with prior research indicating that fast-acting compression leads to reduced speech intelligibility (Stone & Moore, 2004; Drullman, Festen, & Plomp, 1994).

Stone and Moore (2004) also investigated the relative importance of two possible causes responsible for the decreased performance: ”comodulation” and ”modulation reduction”. In a latter paper they also used the term across-source modulation correlation or coherence instead of comodulation to emphasize that this effect of compression introduces patterns of partially correlation from sources that were previously independent (Stone & Moore, 2007). When there are peaks in one signal, the gain applied to all signals in the mixture decreases. This results in signals that were previously independently amplitude modulated to acquire a common modulation component from the compression, reducing their independence and making them more likely to fuse perceptually. This comodulation may lead to a perceptual fusion of attended and ignored speech and thus have disruptive effects related to auditory grouping. In other words, if two or more speech signals with similar comodulation are presented simultaneously, they may interfere with each other and make it more difficult to separate either one (Bregman, Abramson, Doehring, & Darwin, 1985; Hall III & Grose, 1990; Bregman, 1994). Importantly, comodulation should only appear if the mixture of both attended and ignored signals is compressed. We used amplitude modulation on each stream separately, and thus our results should not be driven by comodulation of both streams (Stone & Moore, 2004).

In contrast to comodulation, modulation reduction is important for our case. This is because the gain of the dynamic range compressor used in our loudness matching pipeline is controlled by the most intense components of the input signal, which typically correspond to peaks in the envelope. Fast amplitude compression, which is used in our pipeline, primarily acts on these peaks, reducing their gain. Additionally, our pipeline enhances low intensity signals compared to the uncompressed signal, resulting in a smaller amplitude modulation depth of compressed speech compared to uncompressed speech. As a consequence the fidelity of the envelope is distorted in shape (Stone & Moore, 1992). The non-instantaneous operation of the compressor can cause overshoots and undershoots, which can also contributes to distortion of the envelope (Verschuure et al., 1996).

Since high intensity parts of a speech signal are typically associated with vowels, while low intensity parts are associated with consonants, loudness matched compressed speech will attenuate vowels while enhancing consonants. This results in a distortion of the speech signal and a decrease in the ratio between vowels and consonants. Multiple studies have shown that fast amplitude compression can reduce amplitude modulation depth and intensity contrast (Plomp, 1988; Moore, 1990, 2003; Dillon, 1996). This is crucial for speech recognition as it primarily relies on the temporal cues of speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Stone & Moore, 2004; Peelle & Davis, 2012). Furthermore, both speech streams are subject to reduced amplitude modulation, even though they are not comodulated. This means that they share a similar low amplitude modulation ratio, which can make it more difficult to separate the two streams (Grimault, Bacon, & Micheyl, 2002).

4.8.2 Amplitude compression on ignore speech may facilitate inhibition

Interestingly, we observed a main effect of compression on response speed, which was driven by an interaction with attention. Specifically, normal hearing participants made false alarms with slower responses when ignoring the compressed stream compared to the uncompressed stream. In most studies, researchers investigate response times to target stimuli under different levels of distraction and then compare them. Typically, increasing numbers of distractors increase interference and attentional demand, resulting in delayed response times (e.g., Mazza, Turatto, & Caramazza, 2009). However, in our study, we found no difference in response time between attending to the compressed versus uncompressed stream. The interpretation of slower response times emerging from the attentional background is not well established in the literature. However, one could assume that responding quickly to irrelevant distractors in the ignore uncompressed streams may reflect a failure to inhibit those distractors, while slower responses to the ignore compressed streams could reflect more successful inhibition. It's possible that the increased inhibition of repeats in the compressed stream is due to the aspects of amplitude compression mentioned earlier. Unfortunately, we were limited in the number of false alarms and could not further investigate them according to attentional pairs.

4.8.3 Unexpected lack of effects of amplitude compression applied to single speech streams

Unexpectedly, we observed no significant differences in performance and response speed when either the attended or ignored stream was compressed compared to both streams uncompressed. For the attended stream, we would have expected a decrease in performance when it is compressed, as the decreased modulation depth of the attended stream envelope peaks is associated with reduced perception (Drullman, 1995). However, the greater contrast in amplitude modulation depth between attended compressed and ignored uncompressed could also facilitate stream segregation and lead to compensatory effects (Grimault et al., 2002).

In contrast, we would have expected an increase in performance and faster response times when only the ignored stream is compressed. The decreased amplitude modulation depth in the ignored stream should make it easier to ignore, while the increased contrast in amplitude modulation depth between the attended uncompressed and ignored compressed speech should facilitate selective attention.

One possible explanation for our unexpected result is that the compression applied in our study may not have been strong enough to induce a significant effect, even though we used a fast-acting compressor with a comparable large compression ratio. Future studies could assess this by testing different compression ratios or also varying attack and release times of the dynamic range compressor.

Our task involved detecting short repeats in continuous speech, which is a relatively novel task that likely engages phonological processing more than higher-level semantic processing (Marinato & Baldauf, 2019). Moreover, participants performed relatively well in the task, leaving little room for improvement. Future studies could employ adaptive procedures to track performance and capture potential benefits of compressing the ignored stream. Additionally, it would be valuable to explore behavioral paradigms that assess speech intelligibility, quality, and semantic processing.

4.8.4 Effects of amplitude compression on performance for hearing impaired participants

In general, hearing-impaired participants show similar trends in the data as normal hearing participants, with impaired responses to compressed speech. It seems that, in contrast to normal hearing controls, compression particularly decreases perceptual performance and leads to slower responses to attend compressed speech. Again, caution should be exercised when interpreting the results of this study, given the small sample size of N=7 participants.

The previously discussed points for normal hearing participants also apply to hearing-impaired participants, as they relate to the presented stimulus material. However, due to differences in hearing abilities, one could expect that the compression manipulation would have an even stronger deleterious effect on amplitude-compressed speech for individuals with presbycusis. People suffering from presbycusis have reduced frequency selectivity (Pick et al., 1977; Glasberg & Moore, 1986) and may be relatively insensitive to temporal fine structure cues (Rosen & Fourcin, 1986; Moore & Moore, 2003), but have a good ability to use temporal envelope cues (Bacon & Gleitman, 1992; Turner et al., 1995). Therefore, any form of signal processing that affects the use of temporal envelope cues might be expected to have in particular a negative effect on performance.

4.8.5 Amplitude compression decreases neural tracking of speech

To the best of our knowledge, this is the first study to investigate the effect of amplitude compression on neural tracking of speech in a competing continuous speech paradigm using electroencephalography (EEG) recordings, in both normal hearing and hearing-impaired controls. Our results showed a main effect of amplitude compression on neural tracking in normal hearing participants.

Neural speech tracking relies on the temporal envelope of speech, as has been previously shown in various studies (Ding & Simon, 2012; Kerlin et al., 2010; Mesgarani & Chang, 2012; Rosen, 1992; Golumbic et al., 2013; Etard & Reichenbach, 2019; Kadir, Kaza, Weissbart, & Reichenbach, 2019; Peelle & Davis, 2012; Obleser & Kayser, 2019). Therefore, it was expected that dynamic range compression, a form of signal processing that directly affects the temporal envelope of speech, might impair neural speech tracking. As described in more detail in the section above, the fidelity of the temporal envelope is impaired (vowel to consonant ratio) due to reduction in amplitude modulation depth and overshoot and undershoots in amplitude compressed speech (Stone & Moore, 1992; Verschuure et al., 1996; Stone & Moore, 2007).

The mathematically model to obtain TRF and neural tracking (encoding/decoding accuracy) is based on system identification (Marmarelis, 2004; Ringach & Shapley, 2004). One way to obtain the impulse response function (fully characterization of the LTI system) is by sending a Dirac impulse (which is infinite short in time and high in amplitude but cannot be physically realized) into the system. One other way is to compute a cross-correlation between the measured input and output of the signal. The latter approach is the basis of the mTRF-toolbox (Crosse et al., 2016).

An impulsive input signal contains a wide range of frequencies that can excite the system's resonant frequencies and provide more information about its dynamics (Madisetti, 2018). While the human brain is not linear or time-invariant, studies have shown that the human auditory cortex is highly responsive to changes in the temporal envelope of speech (Howard & Poeppel, 2010). In the context of neural speech tracking, it has been shown that onset envelopes produce the strongest neural tracking and TRF components with the highest similarity to classical ERP (Fiedler et al., 2019; Chalas et al., 2023). Dynamic range compression is a signal processing technique that reduces the gain for high-level signal parts (peaks) and make the signal less impulsive. It also limits the rate of changes in the onset envelope. In summary, the elements described are likely to result in the signal processing driven reduced neural tracking of amplitude compressed speech.

4.8.6 Neural tracking and selective attention: Effects of compression and attentional pairs

In the context of selective attention, our results showed a main effect of attention on neural tracking, with larger tracking of the attended speech compared to the ignored speech. This is in line with several previous studies that have shown enhanced neural responses to attended speech compared to ignored speech (e.g., Mesgarani & Chang, 2012; Golumbic et al., 2013). We did not observe a significant interaction between compression and attention on neural tracking, suggesting that compression did not modulate the effect of selective attention on neural tracking in the first place (Figure 17).

However, due to the quasi-factorial nature of our experimental paradigm, with attended and ignored speech streams presented simultaneously, we investigated the attentional pairs in more detail Figure (19). Based on our hypothesis that applying compression only to the ignored stream would increase neural separation between attended and ignored streams. We expected that neural tracking would be decreased for the ignored stream but also that the neural tracking of the attended stream would be increased even when both streams were unprocessed in this comparison. Since the comparison between attentional pairs in which both streams were presented unprocessed to pairs where only the ignored stream was compressed were in particular interesting for us (19 A). We found a significant difference in neural tracking between the compressed and uncompressed ignored streams, but no significant difference in tracking between the uncompressed attended streams.

This result supports the assumption that compression reduces the neural processing of speech, which in turn decreases neural tracking. The main driver of this effect is likely the distortion of the fidelity of the envelope, as discussed earlier. This is further supported, since remaining attentional pair comparisons indicate significant differences between compressed and uncompressed streams but no signs of interaction effects (19 B-F).

However, neural tracking of speech reflects most likely both the acoustic information of sound as well as top-down attentional processes and thus might be a correlate of both auditory object formation and selection (B. Shinn-Cunningham et al., 2017). In attentional pairs where both a compressed and uncompressed stream are presented, the contrast in the modulation depth between the two signals is larger, resulting in more distinct attentional cues. This difference could potentially facilitate stream selection (Grimault et al., 2002; Bregman, 1978; B. G. Shinn-Cunningham, 2008). Hence, the observed difference between ignored streams in neural tracking could also be influenced by listeners' ability to actively ignore distracting speech more easily by suppressing neural responses to it (Schneider et al., 2022; Wöstmann, Alavash, & Obleser, 2019; Fiedler et al., 2019).

To investigate the contribution of top-down attentional processes such as active ignoring and the impact of signal processing (i.e., distorted envelope due to compression) on the difference in ignored neural tracking between attentional pairs, it would be useful to consider the temporal resolution of neural tracking and temporal response functions (TRFs). TRFs are often interpreted in a similar way to auditory evoked potentials (AEPs), with each component representing neuronal activity along the auditory pathway (T. W. Picton et al., 1974). Differences in the latencies of TRF components can provide information about the underlying neuronal origins of those components (for review, see Brodbeck & Simon, 2020). Fiedler et al. (2019) found a late TRF-N2 component associated with active ignoring under varying signal-to-noise ratio conditions. However, in our study, we found no significant differences between the ignored streams of the compared attentional pairs when comparing TRFs and time-resolved neural tracking. One possible reason for this could be that the sensitivity of these methods is too low to detect subtle differences. Future studies using decoding-based analyses, more EEG channels or different compression parameters could be more sensitive in detecting potential differences in time-resolved analysis between attentional pairs under amplitude compression.

4.8.7 Hearing impaired participant showed larger neural tracking responses compared to normal hearing participants

Our findings are consistent with some previous literature, which has also reported stronger overall neural tracking in individuals with age-related hearing loss (Figure 22 A). For example, Petersen et al. (2017) reported larger neural tracking of the ignored stream for hearing impaired listeners, and both Decruy et al. (2019) and Fuglsang et al. (2020) found that hearing impaired listeners show larger neural tracking responses to target speech. In quiet environments, hearing-impaired listeners exhibited increased neural speech tracking, along with delayed neural responses that had longer latency compared to age-matched controls (Gillis et al., 2022). Furthermore, Schmitt et al. (2022) even reported enhanced neural speech tracking with increasing hearing loss. This finding is in line with previous literature that suggests hearing-impaired individuals rely even more on temporal envelope cues as a possible compensation for their reduced processing of the temporal fine structure (Bacon & Gleitman, 1992; Turner et al., 1995; Rosen & Fourcin, 1986).

On the other hand, there are also opposing views regarding "the stronger, the better". In ageing research, the enhanced amplitude of sensory evoked responses is often associated with listeners deficits in inhibitory control (e.g., Alain, Roye, & Salloum, 2014; Presacco, Simon, & Anderson, 2016). In addition, higher cognitive processes were associated with the cortical representation of the speech signal. Higher neural tracking could also be related to an inefficient use of cognitive resources and a decrease in cortical network connectivity (Peelle, Troiani, Wingfield, & Grossman, 2010).

However, other studies have found conflicting results; for instance, Tune et al. (2021) found no increased neural tracking with hearing loss or age, and Gillis et al. (2023) reported that acoustic speech and linguistic representational neural tracking decreased with age. While there are discrepancies in the literature, our results align with previous findings of stronger neural tracking in individuals with age-related hearing loss.

The main effect of amplitude compression on neural tracking appears to be similar for both normal hearing and hearing-impaired participants. Specifically, amplitude compression decreases neural speech tracking in both groups. Therefore, the general conclusions drawn for normal hearing participants can also be applied to hearing aid users.

However, when we compared the neural tracking of attend-compressed and ignore-uncompressed versus attend-uncompressed and ignore-compressed speech within normal hearing participants, we found a tendency that the to-be-attended stream is stronger tracked when the ignored stream is compressed (Figure 25 A). There is the possibility that manipulating the unattended stream also influences the processing of the attended stream. (Makov & Zion Golumbic, 2020) reported that manipulating the rhythmic regularity of distracting tones affected not only distractor but also target tones neural processing. They emphasise the possibility that suppressing the target can result in changes in how the target is processed.

Comparing these differences in hearing impaired and normal hearing participants, we found that hearing impaired participants show larger differential tracking between both the attended and ignored streams (Figure 25 B). Normal hearing participants performed very well in the task and perhaps the compression manipulation on the ignored stream was insufficient to activate the need for distractor suppression or target enhancement. In contrast, hearing impaired participants had the tendency to overall perform worse and thus may profit more from an ignored compressed talker. The comparison between NH and HI of the compressed versus uncompressed ignored streams revealed a consistent weaker tracking of the HI participants that may indicate an additional suppression of the compressed stream that may facilitate the processing of the attended stream as indicated by the increase in neural tracking.However, to really distinguish cleanly between target enhancement and distractor suppression, an additional baseline would be very supportive(Wöstmann et al., 2022).

Furthermore, recent studies have reported that better speech comprehension is associated with enhanced neural tracking in hearing-impaired individuals (Schmitt et al., 2022). Additionally, there is evidence that neural speech tracking correlates with speech intelligibility (Peelle et al., 2013) and behavioral indices of speech comprehension (Etard & Reichenbach, 2019). However, we found no behavioral evidence based on the detection of repeats that hearing-impaired participants also performed better here, which suggests that future studies are needed to explore the relationship between neural responses and speech comprehension or quality.

The results of this study provide new insights into the neural and behavioral effects of compression on attended and ignored streams in normal and hearing impaired individuals. Specifically, the study reveals that compression on both attended and ignored streams leads to decreased behavioral performance and decreased neural speech tracking. Additionally, the study demonstrates that the largest neural separation is observed when only the ignored stream is compressed in hearing impaired participants, highlighting the potential benefits of compression for individuals with hearing impairments.

Overall, this research advances our understanding of the complex interplay between attention and compression in auditory processing and provides important foundation for developing effective interventions for individuals with hearing impairments.

4.9 Limitations

To avoid confounds with loudness perception, we used a loudness matching procedure to match the perceived loudness of uncompressed and compressed speech. However, it should be noted that the peaks in the envelope of the compressed speech are still reduced compared to the uncompressed speech, and low-level signals are partly enhanced in the compressed speech to match the loudness between the two versions. This means that there may be some energetic masking effects, even with the two speech streams almost uncorrelated (Brungart, 2001).

In attentional pairs where one stream was compressed and the other was not, this could mean that the attend compressed stream is more likely to mask the ignore uncompressed stream for low-intensity parts, and vice versa for high-intensity parts. Similarly, in attentional pairs where the attend stream is uncompressed and the ignored stream is compressed, the ignored stream is more likely to mask the attended stream for low-intensity parts, and vice versa for high-intensity parts. In other words, loudness-matched compressed speech is more likely to mask the uncompressed speech at low levels and vice versa for high-intensity parts.

Although both streams are uncorrelated as well as speech streams can be, it could be potentially beneficial to investigate this in more detail. The low intensity masking might work to some degree in favor of the loudness matched compressed speech since on average low-intensity signals are more prominent in the compressed stream. Low-level parts of a signal are associated with consonants that also play an important role in speech intelligibility (e.g., Villchur, 1973). Based on this, there could be a compensatory effect between the low-level masking of uncompressed and the high-level masking of compressed speech. Future studies are needed to investigate this relationship in more detail. For instance, one could add different signal-to-noise ratios (SNRs) to investigate potential interactions with SNR and compression.

Another limitation of the present study is that we did not perform a brain-behavior relationship analysis to directly link changes in neural processing to changes in behavioral performance. An analysis of the relationship between our behavioral and neural data would reveal if there is a connection between the two, providing additional insights into the neural mechanisms underlying selective attention and compression. Additionally, while the current study included both normal hearing and hearing impaired participants, the sample size for the hearing impaired group was relatively small (n=7) and thus my not be representative of the population. However, it is important to note that the hearing loss was similar in terms of slope and PTA among hearing impaired participants (see Figure 12). Moreover, a previous study (Verschueren, Vanthornhout, & Francart, 2021) reported that neural speech tracking is robust to stimulus intensity, and the attention and compression effects were consistent (see Figure 22) despite the small sample size. In addition, the group of hearing-impaired participants can be considered a sub-group of the normal-hearing participants and may be analyzed within the same statistical approach. Nevertheless, future studies with larger sample sizes and a mixed model analysis incorporating covariates such as overall presentation level, age, and degree of hearing loss would be beneficial to further explore the effects of attention and compression on neural processing and behavioral outcomes in this population.

4.10 Conclusion

The study sheds light on the intricate interplay between attention and compression in neural processing and behavior, deepening our understanding of how amplitude compression impacts the neurophysiological mechanisms underlying selective auditory attention during ongoing speech. This study potentially paves the way for novel hearing aid algorithms. However, the possibilities of amplitude compression are not fully explored, and future research could explore multi-band compression, comodulated compression of multiple sound sources in the background, and sidechain compression that compresses the ignored based on the attended stream. These possibilities are further explained in Section 5.6 of this thesis.

5 General discussion

The present thesis investigates behavioural and neural signatures of selective attention to speech. The goal of this thesis was twofold: first, the top-down contributions of target enhancement and distractor suppression to selective attention were investigated by implementing a neutral control condition to separate those sub-processes utilising a novel psychophysically augmented continuous speech paradigm and electroencephalography (EEG). Second, how far might mechanisms of selective attention interact with amplitude compression. In particular, we investigated how normal-hearing listeners and older adults suffering from presbycusis cope with amplitude compressed speech in a multi-talker situation. To this end, we conducted a pilot study on neural responses to different compression ratios, modelled the peripheral fate of amplitude compressed speech, performed and evaluated loudness matching between amplitude compressed and uncompressed speech, and measured behavioural and neural responses from normal and hearing participants to amplitude compressed speech in a psychophysically augmented continuous speech paradigm. The discussion then concentrates on the main findings after summarising the experimental results. If there are more detailed discussions about specific findings, these can be found in the respective experimental chapters.

5.1 Summary of experimental results

Study 1 found that selective attention is implemented by the enhancement of target speech, rather than the suppression of distraction. The results showed that listeners committed more false alarms originating from the distractor speech than the neutral stream. However, the neural representation of target speech was enhanced, while no suppression of distraction was observed below the neutral baseline. These findings suggest that target enhancement is the primary mechanism underlying selective attention.

Study 2 presented a narrative story with different compression and expansion ratios (1:2 and 1:8 for compression and 2:1 for expansion) to six individuals with normal hearing, as well as an unprocessed baseline condition. The main objective of the pilot experiment was to determine an appropriate compression ratio to be used in a follow-up study. A 1:8 compression ratio significantly reduced the brain's ability to track speech compared to unprocessed speech. This was shown by a significant decrease in decoding accuracy for the 1:8 compression ratio, which was observed in all participants. The 1:8 ratio was used in the follow-up experiments.

Study 3 investigated the effects of amplitude compression and selective attention on neural separation and behavioural response in normal-hearing participants using a continuous speech paradigm. The results showed that compression on both attended and ignored streams decreased behavioural performance and neural speech tracking and increased neural separation when only the ignored stream was compressed.

Study 4 (simulation) used a computational model of the human auditory periphery to simulate model outputs to (compressed) speech signals, modelling the firing rate of the auditory nerve and the envelope following response for both normal hearing and hearing-impaired participants. The results indicated that hearing status and frequency significantly impacted the output of the auditory nerve (AN), with higher spike rates in the normal hearing group and higher spike rates for higher frequencies. Importantly, there was no significant effect of signal manipulation on the firing rate, and if anything, the amplitude was increased for loudness-matched compressed signals. The simulation results suggest that changes in the auditory periphery did not confound the effects observed in the neural speech tracking study and the follow-up study for hearing impaired participants.

Study 5 aimed to study the impact of amplitude compression and selective attention on neural separation and behavioural response in individuals with hearing impairment. The findings demonstrated comparable patterns to those observed in normal-hearing participants, with reduced performance and neural tracking for amplitude-compressed speech. However, unlike the normal-hearing participants, the results showed enhanced neural speech tracking for the attended stream when only the ignored stream was compressed.

5.2 Terminological considerations

Before discussing the general implications of both parts, it's important to consider the terminological choices used in this thesis. Terminology can lead to conflicting theoretical inferences, as discussed more recently by (Makov, Pinto, Yahav, Miller, & Golumbic, 2023). In close analogy to (Seidl et al., 2012), the terms ”target”, ”distractor”, and ”neutral” were used within this thesis and can be considered as theory-derived terms without further context. These terms are rather abstract and generalise at a high level. Furthermore, they are likely to make assumptions about underlying cognitive processes or be used differently in different studies. On the other hand, methodological-based terms are advantageous because they do not require making assumptions about the inner state or cognitive operations (Makov et al., 2023).

However, within this thesis, the terms ”target”, ”distractor”, and ”neutral” are used as placeholders for task-relevant, task-irrelevant, and never-task-relevant, which are methodologically based terms. This trisection was used especially to separate the terms that occur in the attentional background (i.e., distractor and neutral). If one is precise, the terms attentional background and foreground are also theory-driven terms, as one would assume that the participant's attentional operations are associated with their separation.

The term neutral plays a special role here, as it refers to the never-task-relevant stream and is used as a control or baseline to separate target enhancement and distractor suppression. At first glance, one might expect the neutral stimulus to be a stimulus that does not elicit a natural response. However, this is not the case for an additional speech stream in a cocktail scenario, as described in greater detail in Section 3.

The second part of the thesis moves away from the neutral stream and employs the terms ”attended” and ”unattended” (sometimes also referred to as ”ignored”) streams, as previous studies have done (e.g., Ding et al., 2012). However, these terms are theory-derived, as they assume that participants truly allocate attention to the cued stream. Therefore, more precise methodological terms might be cued and uncued streams. In general, selecting appropriate terms to describe auditory attention based on previous studies or categories such as theory or methodological terms is not trivial. However, increased awareness of this issue will be beneficial for future studies.

5.3 The role of short repeats as a behavioural measure for assessing selective attention in continuous speech

Continuous speech is often preferred over trial-based designs in experiments, particularly when investigating speech perception in multi-talker situations, as it provides increased ecological validity (Hamilton & Huth, 2020). However, one significant disadvantage is the limited availability of behavioural data. But behavioural measures and their relation to cortical speech tracking are key to investigating its meaning. Typically, comprehension questions are asked infrequently or after the experiment, making it challenging to determine the relevance of neural responses to the task. This creates a difficulty in studying brain-behavior relationships as neural recordings are fine-sampled while comprehension questions are rather discrete. To address this issue, short, repeated segments of speech were included in the speech streams in this study, which required participants to detect them quickly in the target stream and ignore them in the attentional background (Marinato & Baldauf, 2019). This allowed the measurement of response times and hit and false alarm rates for the repeats in different speech streams.

One question that arises here is at what processing level the repeats are processed. We included quasi-randomly (some constraints; see Section 3 for more details) repeated segments of the speech stream with a length of 400 ms in the speech streams. One interesting characteristic of the repeated segment is that it has the same acoustic properties as the preceding segment before the repetition. Therefore, detecting the repetition based on low-level acoustic features such as changes in intensity or frequency would be unlikely, which would be the case if, for example, the names of the participants were randomly included in the streams. Our findings support this, as repeats in the neutral and distractor streams did not evoke a significant temporal response function in contrast to the repeats in the target stream. Thus, higher-level processing of the repeats, such as grouping based on phonemes and syllables, is likely involved in repeat detection (Marinato & Baldauf, 2019).

The duration of spoken syllables in German depends on several factors, such as the number and type of sounds that make up the syllable, and their respective durations, the speaking rate or tempo of the speaker, and the surrounding context in which the syllable occurs. Generally, the longer a word is, the shorter its syllables. For instance, the average duration of spoken syllables in German is 388.51 ms for a word containing 1 syllable and 172.54 ms for a word containing 7 syllables, and 1 syllable contains on average 2-3 phonemes, the smallest unit of speech (Altmann & Schwibbe, 1989; Sievers, 1876). Therefore, in a range of 400 ms fall on average after roughly 2 syllables and 5 phonemes. In conclusion, the participants most likely processed the speech stream at the phonetic level of streaming speech based on grouping phonemes and syllables to recognise the repeat in speech and to respond to it.

Some of the repeats could thus also include shorter word and are thus also processed and the verly low level of morphemes word meaning (e.g., Di Sciullo & Williams, 1987). However, higher levels of semantic processing such as sentence semantics and beyond the sentence boundary, are probably not covered by the repeat detection task. But the listener's goal is typically to understand their interlocutor at a higher semantic level, in order to have a productive conversation. This is one limitation of the repeat detection task.

One could even go a step further and consider the repeat detection task and attending to the content of the audiobooks as a dual-task. Are these two tasks completely independent? Could participants be solely focused on detecting the repeats without processing the content of the audiobooks? We also asked comprehension questions regarding the content of all streams at the end of the experiment. Due to the nature of the paradigm, which involved fast switching of attention between two streams, it was not always clear which stream the answers pertained to. Additionally, since the questions were asked at the end of the experiment, there was a memory component involved. However, results indicated that participants processed the streams at higher semantic levels. More importantly, our brain-behavior analysis revealed a significant relationship between the neural tracking of continuous speech and repeat detection performance. In other words, the larger the neural tracking for the target stream, the better the repeat detection performance. These findings support a relationship between the long-term tracking of speech and the repeat detection task and demonstrate the feasibility of our new continuous speech paradigm.

The role of repeats in the attentional background of speech streams is unique. As mentioned earlier, the repeated segment shares the same acoustic properties as the preceding segment before the repetition, making it unlikely for the repeats to pop out in the attentional background. Hence, it is possible that the streams in the attentional background have to be preattentively segregated to some extent to have the potential to trigger a false alarm. This relates to the ongoing discussion on whether the process of separating auditory objects is preattentive or whether it is influenced by attention, and would rather add to the former (Carlyon, 2004; B. G. Shinn-Cunningham, 2008; Puvvada & Simon, 2017). Additionally, we observed a significant difference in the rate of false alarms between the distractor and neutral streams, despite both being considered part of the attentional background. This finding challenges the assumption that background sources are unsegregated and suggests that there may be some degree of preattentive segregation even in the background. However, we observed only a relatively low number of false alarms, and we quasi-randomized their occurrence, which also randomised the acoustic features and processing level associated with them. Therefore, it is unclear if there is some form of clustering in the repeats based on their features that is more likely to create false alarms. Further research is necessary to investigate this issue. It is also unclear whether participants ”detected” a repeat in the background and successfully suppressed it or simply did not ”detect” the repeat, which both would result in no response to repeats in the attentional background.

While the repeat detection task in continuous speech has some drawbacks, such as its limited ability to reflect high-level semantic processing in a multi-talker paradigm, it offers several advantages over discrete comprehension questions. For instance, it provides a much more fine-resolved behavioural measure, and yields richer behavioural data that can be analysed using signal detection theory (e.g., hits, false alarms) and response times to the repeats. It should be noted, however, that the repeats in continuous speech may reduce their ecological validity. On the other hand, the more continuous nature of the task may engage participants more, while the hit-false alarm ratio is a good indicator of whether the participant is allocating attention as specified for the task, an important factor for neural analysis. Finally, our research has shown a relationship between the long-term neural tracking of speech and performance on the repeat detection task. In sum, the repeat detection task has some advantages but also some disadvantages and is probably not yet the final answer to unravel the precise relationship between brain and behaviour in understanding cognitive processes related to selective attention in multi-talker situations.

5.4 Implication of the psychophysically augmented continuous speech paradigm on stream formation

In the first part of the thesis, we investigated the mechanisms of selective attention based on top-down attention, specifically attention switching (negative priming) between two task-relevant streams while also having a task-irrelevant stream in the attentional background.

We demonstrated that the neural representation of target speech is specific to processes of attentional gain for behaviorally relevant target speech rather than neural suppression of distraction. We quantified target enhancement as an increased cortical gain reflected in the neural tracking of the target stream versus the neutral stream, and we quantified distraction suppression as a decreased neural tracking of the distractor stream versus the neutral stream.

One could argue that a prerequisite for this is that all three streams are represented as different auditory objects. It is widely accepted that an auditory scene is perceived in terms of auditory objects (Bregman, 1978, 1994; Griffiths & Warren, 2004; Shamma et al., 2011; B. G. Shinn-Cunningham, 2008) and some believe that suppression operates on the representation of objects (Geng, 2014; Noonan et al., 2018; Daly & Pitt, 2021).

However, this touches on the long-standing debate regarding at what level of hierarchy auditory object segregation is implemented: preattentive or actively influenced by selective attention (Carlyon, 2004; B. G. Shinn-Cunningham, 2008; B. Shinn-Cunningham et al., 2017; Shamma et al., 2011). There is mixed evidence about spectro-temporal-based, acoustic-based, and object-based representations of the auditory scene in the core auditory cortex.

It has been shown that core neural activity in the auditory cortex reflects acoustic characteristics of speech, such as spectro-temporal features (Ding & Simon, 2013; Okada et al., 2010). On the other hand, it has been proposed that neural auditory objects are formed, and that representations of dynamic sounds are influenced by task demands in early auditory cortex (Nelken & Bar-Yosef, 2008; J. Fritz, Shamma, Elhilali, & Klein, 2003).

More recently, Puvvada and Simon (2017) used a three-speaker paradigm to investigate the cortical representation of speech in an auditory scene. They found no distinct representation between the two streams in the attentional background, even at higher-order auditory areas, and the mix of unattended streams was more faithfully represented than the separate representations of both unattended streams. These results suggest that speech streams in the attentional background are not represented as distinct auditory objects but rather as one merged auditory object resulting from the mix of unattended streams.

In the context of our study, these results would imply that distractor suppression would not be measurable by contrasting two speech streams in the attentional background since both are represented as one object. At first glance, our neural results are broadly in line with this conclusion, but note that Puvvada and Simon (2017) had not applied any differential task manipulation. to the two background speech streams, which we aimed to achieve here.

We made three important changes. First, (Puvvada & Simon, 2017) investigated the mix of the three streams, resulting in single channel representation without spatial separation. We used three spatially separated streams, thus participants could make use of spatial cues that likely influence auditory stream segregation (Darwin & Hukin, 2000). Secondly, we manipulated the attentional streams differently by task. When attentional switching was indicated by a spatial cue, the two streams alternately represented the attended object. In other words, both streams were likely to form an object (at least for cued trials) regardless of whether the formation of auditory objects occurred preattentively or was modulated by attention. We hypothesised that this competition in the neural representation of the task-relevant streams is associated with inhibition, retrieval mechanisms (Tipper, 1985; Frings et al., 2015), and the representation of event files (Hommel, 1998). Third, we incorporated the repeat detection paradigm (Marinato & Baldauf, 2019) into a continuous speech paradigm.

In contrast to the neural tracking results, the behavioural results suggest distinct processing of the attentional background, as discussed in more detail in the previous section. The shared acoustic properties between a repeated segment and its preceding segment make it difficult for the repetition to be detected in unattended streams without preattentive segregation of the auditory streams. Additionally, we observed a significant difference in the rate of false alarms between the distractor, which also suggests a separate processing of the attentional background. But what is the relationship between the neural speech tracking and behavioral results, especially regarding the attentional background? We have shown that participants' performance in the repeat detection task is related to the neural tracking of the target stream. We found no evidence for a relation between stream and repeat tracking of the neutral stream and participants' performance, and also no statistical evidence that stream tracking of the distractor stream is related to perceptual performance. But a tendency ($Z_{Wald} = -1.44$, $p = 0.147$) that lower neural tracking of distractor speech is associated with increased performance. This may also be an indication of a distinct relation between the neutral stream and ignored stream on behavioural performance.

In sum, the results are rather mixed relating to the more general questions of whether streams are formed preattentively and whether multiple streams can coexist, or whether attention is required to form a stream. Additional research is needed to further explore these issues.

5.5 Investigating the sub-mechanisms of selective attention: Intersection of findings from two parts of the thesis

In the first part of this thesis, we examined the mechanisms of top-down selective attention. We presented speech streams on different spatial positions and narrative stories spoken by different talkers. Thus, listeners' most likely used spatial and frequency cues to separate speech streams. The assignment of speakers to locations was randomised across participants, and no acoustic features of the speech streams were manipulated.

In contrast to the first part of the thesis, in the second part we used dynamic range compression as an acoustic manipulation to speech streams. However, we did not include a neutral control condition. We made this decision for two main reasons. First, based on the neural speech tracking results, we did not find any support for distractor suppression or differential processing of speech streams in the attentional background. Second, we aimed to mimic hearing aid processing by presenting speech streams in both the front and back of the participant.

In the first and second parts of this thesis, we learn about the sub-processes of selective attention, target enhancement, and distractor suppression. When no acoustical manipulation was applied to the speech streams in the first part, evidence for target enhancement was found, but not for distractor suppression. However, classical attention theories assume the possibility of some form of distractor suppression, where the brain selectively filters out distracting information. It is possible that different neural mechanisms can implement distractor suppression, which could explain the absence of evidence in the first part (Broadbent, 1958; Treisman, 1960; Gaspelin & Luck, 2018; Wang & Theeuwes, 2018a; Noonan et al., 2018; Wöstmann et al., 2022; Daly & Pitt, 2021).

There are two types of distraction suppression: proactive and reactive. Proactive suppression is processing done before the distraction appears, while reactive suppression is processing done after the distraction has drawn attention. Alpha power is usually associated with proactive suppression, while the characteristics of neural tracking of speech reflect reactive distractor suppression. However, the study found that reactive suppression is absent in auditory cortex responses in a multi-talker situation. Thus, future studies are needed to investigate proactive suppression using other neural measures (Geng, 2014; Wöstmann, Alavash, & Obleser, 2019; Noonan et al., 2018).

Participants performed well on the task, as indicated by high hit rates and generally low false alarm rates. Our results suggest that this performance was mainly related to the neural speech tracking of the target stream, which enhanced processing of the target. However, it is possible that the task was too easy, leading to a lack of engagement of suppressive mechanisms. To increase the difficulty of the task and activate suppressive mechanisms, one approach could be to manipulate the acoustics, such as by varying the signal-to-noise ratio (SNR) between attended and ignored signals. For example, a study by Fiedler et al. (2019) found that, under adverse listening conditions, such as manipulations of the SNR led to the activation of suppressive mechanisms.

In the second part of the thesis, we manipulated the acoustics in terms of dynamic range compression. We found some indications of suppressive mechanisms in neural speech tracking. Dynamic range compression on ignored speech increased neural tracking of attended speech in hearing-impaired participants, suggesting that suppression (possibly facilitated via compression) influences the processing of the attended stream (Makov & Zion Golumbic, 2020; Daly & Pitt, 2021). Furthermore, there is evidence that the potential influence of target and distractor suppression relies on separable underlying mechanisms, as indicated by different activity patterns (Jaeger, Bleichner, Bauer, Mirkovic, & Debener, 2018). However, a baseline control condition (similar to the one in study 1) is needed to reliably distinguish between this mechanisms: target enhancement and distractor suppression (Wöstmann et al., 2022).

In sum, one important intersection of the first and second parts of this thesis is the investigation of the sub-processes of selective attention. While the first part found evidence for target enhancement but not for distractor suppression, the second part showed some indication of suppressive mechanisms in neural speech tracking. The second part demonstrated that dynamic range compression on ignored speech could increase neural tracking of attended speech, suggesting a possible suppression mechanism that influences the processing of the attended stream. These findings indicate the complexity of selective attention and the possibility of different mechanisms involved in target enhancement and distractor suppression. Further research is needed to better understand the neural mechanisms underlying these sub-processes of selective attention.

5.6 Implications on hearing aids and future research

To date, amplitude compression or dynamic range compression is widely used in most modern hearing aids. The primary goal is to compensate for loudness, and the compressor parameters vary depending on the listening situation, often being applied in a frequency-dependent manner. However, the pros and cons of amplitude compression are still under discussion. (Braida et al., 1979; Dillon, 1996; Souza, 2002). Previous studies have indicated that hearing aids that compress the total sound scene with the same compression often have negative effects on performance, particularly in multi-talker situations where target speech and distracting speech are compressed together, undergoing a comodulation (Stone & Moore, 2004). More recently, hearing aid algorithms are able, via beam-forming technology, to apply different signal processing techniques,including compression, to separate sound sources from different locations (Jensen et al., 2021).

In the second part of the thesis, we simulated independent spatial signal processing by presenting compressed and uncompressed speech over free-field loudspeakers in a spatial competing talker paradigm. We investigated the interplay of fast-acting amplitude compression with selective attention and found that, in general, behavioural and neural responses were impaired when amplitude compression was applied to speech streams. Particularly, hearing-impaired participants showed reduced speech tracking of the amplitude-compressed streams and increased neural speech tracking of the attended stream when only the ignored stream was compressed, compared to both streams being uncompressed.

Overall, the results support the hypothesis that fast-acting amplitude compression on both streams (even when not comodulated in our experiment) impairs neural speech tracking and performance. Additionally, the study suggests that hearing-impaired participants may benefit from independent compression applied only to ignored sources, as indicated by enhanced neural speech tracking. However, the study did not find increased behavioural performance to support this result. Future studies could investigate the effect of amplitude compression on other behavioural measures, such as speech intelligibility. Moreover, sound quality is also crucial for the acceptance of hearing aids, and further studies could evaluate the effect of compression on quality ratings.

We only scratched the surface of signal processing in terms of compression. Our study applied wideband (single-channel) compression to the stimulus material and investigated TRFs based on the cochlea-filtered broadband envelope. However, compression can alter the envelope shape of an audio signal in a way that adds distortion components to the modulation spectrum that were not present in the original signal (Stone & Moore, 2007). In addition, neural speech tracking not only relies on the speech envelope but also on temporal fine structure (Ding, Chatterjee, & Simon, 2014; Obleser, Herrmann, & Henry, 2012). Research that explicitly considers the frequency domain could provide additional insights. To explore the impact of fast-acting compression to the modulation spectrum on neural tracking, it would be interesting to investigate the spectro-temporal response functions (sTRFs) of amplitude-compressed speech (Drennan & Lalor, 2019; Fiedler et al., 2017; Kraus et al., 2021). Additionally, the use of multi-channel compression on neural speech tracking and performance on independent channels should be further explored. For hearing-impaired participants, carefully set channel-dependent compression is associated with improved speech intelligibility and quality and may support the acceptance of such algorithms (Souza, 2002).

Further research is needed to investigate the role of across-source modulation coherence (Stone & Moore, 2007, 2004) and its impact on selective attention, neural tracking, and performance in multi-talker situations. This would be particularly interesting when multiple objects in the attentional background are comodulated (compression leads to patterns of modulation that are partially correlated across previously independent source) but the target talker is not. This could lead to a perceptual fusion of distractors, which may facilitate stream segregation and selection (Grimault et al., 2002; Shamma et al., 2011), resulting in potential better performance, especially for hearing-impaired listeners in situations with multiple distractions.

Dynamic range compression is a crucial tool in hearing aids, as well as being used in music production. In music production, side chain compression is often used to modify the behaviour of the compressor using a separate audio signal (Oliveira, 1989). One example of this is reducing the volume of a bass guitar signal when a kick drum transient occurs, creating a ”ducking” effect that allows the two instruments to blend better. This same side-chain compression technique could also be a beneficial feature in hearing aids. For instance, in a multi-talker situation where there are several speech sources, a listener may want to focus on one particular talker. The compressor could then reduce the gain of the ignored talker(s) based on the transients in the attended talker's speech signal, which could be controlled and adjusted via the attack and release times. Since speech transients play an important role in speech comprehension (for review, see Peelle & Davis, 2012), reducing the gain of the ignored talker(s) based on the transient could be beneficial for speech comprehension. This technique could help to improve the intelligibility of the attended speech source, while reducing the distracting effect of the ignored talker(s).

5.7 Limitations

It is important to note some limitations of our present work. The limitations of the individual studies are described in the corresponding section. Here, the more generally applicable limitations are described. While our newly introduced behavioural repeat detection task in continuous speech has some advantages, it also has some drawbacks, which we discuss in more detail in section 5.3. The main limits are likely that the task, in general, only led to a small number of false alarms, and that the repeat detection task does not cover high-level semantic processing.

For the former, the repeat detection task has the great advantage that it provides fine-resolved behavioural data from speech streams in the attentional background. However, the total number of false alarms is comparably low. There are probably different ways to increase the false alarm rate. At first glance, a straightforward possibility would be to reduce the length of the repeat. The shorter the repeat, the less information the participant can use to form the repetitions, and the more similar the repeats become between streams, potentially making them more difficult to separate, which may increase the false alarm rates. On the other hand, the shorter the repeats get, the more they reflect low-level acoustics, making them more likely to pop out of the streams. Another possible option would be to make the task more difficult without changing the repeat length, for instance, by adding acoustical manipulations to the task, such as SNR manipulations. However, this must be well thought out and, of course, depends on the respective hypotheses.

For the latter, it is always a trade-off between sampling behaviour at a high rate and speech comprehension. We aimed to obtain more finely resolved behavioural data to have a measure that can compete (at least to some extent) with the comparable, very-high-sampled neural recordings. We found a significant brain-behavior relationship, and indeed, neural tracking of target speech predicts the proportion correct of the repeat detection task. However, the repeat detection task is insufficient for high-level speech comprehension or even extracting meaning from a conversation. Future studies are still needed to address this complex but important issue in the field. One possibility could be to cautiously and well-consideredly combine different sampled behavioural measures that address different levels of processing. In addition to the repeat detection task, one option could be to randomly stop the presentation and ask the participant to repeat the last sentence or a certain number of words (O'Sullivan et al., 2017), or to collect self-reported intelligibility scores (Ding & Simon, 2013).

We claimed that we used a more ecologically valid experimental paradigm in contrast to more trial-based designs. We used narrative stories as stimuli that appear in real-life scenarios but do not reflect typical conversation. They still operate on a continuum between well-controlled and ecologically valid real-life paradigms, perhaps leaning more towards the latter. The additional attention-switching components in our design make the task more interactive, as in challenging listening situations, such as in a bar, one is more likely to change interlocutors. On the other hand, the embedded repeats in the speech stream make the speech streams more unnatural (since repeats do not usually appear in everyday conversation) and move the paradigm on the continuum more towards well-controlled paradigms. In addition, real-life listening scenarios have much more variation in background noise than competing talkers reflect as narrative stories, and participants can usually use visual information, such as lip-reading, to facilitate speech comprehension. Once again, this boils down to a trade-off between ecologically valid paradigms and more trial-based, controlled paradigms. To choose one over the other or even to choose a combination of both, which we claimed within this thesis, needs to be reconsidered for future studies and depends on the research questions asked.

6 Conclusions

In this thesis, we investigated the sub-processes of selective attention in complex auditory environments using psychophysically augmented continuous speech paradigms. The results of the first study demonstrated that selective attention is achieved by enhancing the target speech rather than suppressing the distraction. This result was supported by the brain-behaviour relationship, indicating better performance with increased neural speech tracking of the target streams. This provides valuable insights into the mechanisms of selective attention. This finding challenges current models of enhanced neural responses to speech and emphasises the importance of considering specific sub-processes of selective attention, such as target enhancement, when examining the neural mechanisms underlying speech processing.

The second part of the thesis investigated the effects of dynamic range compression on neural separation and behavioural response in normal hearing and hearing-impaired participants. The results lay a foundation for our understanding of how amplitude compression affects the neurophysiological mechanisms underlying selective auditory attention during ongoing speech, with potential implications for the development of novel hearing aid algorithms. We show that fast-acting compression in general impairs performance and leads to decreased neural speech tracking. On the other hand, applying dynamic range compression only to ignored talkers in a multi-talker situation can lead to increased neural separation between attended and ignored talkers in hearing-impaired listeners. However, we found no associated increase in performance related to repeat detection. Further studies are needed to also investigate the effect of dynamic range compression on additional behavioural measures.

While our study provides important insights into the neural mechanisms underlying selective attention and the interplay with dynamic range compression, there are limitations that should be acknowledged. The repeat detection task provided rich and finely resolved behavioural data in contrast to common methods. On the other hand, it did not cover high-level semantic processing, which is important for speech comprehension. The studies were conducted in controlled laboratory environments with narrative stories as stimuli and may not fully represent real-world listening scenarios. Additionally, the study with hearing-impaired participants was conducted with a relatively small number of participants, and further studies with larger sample sizes are needed to validate the findings.

In conclusion, this thesis contributes to the ongoing debate in attention research by providing new insights into the sub-processes of selective attention in complex auditory environments. Our findings suggest that dynamic range compression can have different effects on behavioural performance and neural speech tracking, and that the enhancement of the target speech is the primary mechanism behind selective attention. Future studies are needed to investigate the neural mechanisms of selective attention and validate our findings in larger samples and real-world listening scenarios.

References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*, *28*(15), 3958–3965.

Ahlfors, S. P., Han, J., Belliveau, J. W., & Hämäläinen, M. S. (2010). Sensitivity of meg and eeg to source orientation. *Brain topography*, *23*, 227–232.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Alain, C., Roye, A., & Salloum, C. (2014). Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Frontiers in systems neuroscience*, *8*, 8.

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology*, *42*(6), 770–779.

Altmann, G., & Schwibbe, M. H. (1989). *Das menzerathsche gesetz in informationsverarbeitenden systemen.* Georg Olms Verlag.

Ananthakrishnan, S., Krishnan, A., & Bartlett, E. (2016). Human frequency following response: neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss. *Ear and hearing*, *37*(2), e91.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in cognitive sciences*, *16*(8), 437–443.

Bacon, S. P., & Gleitman, R. M. (1992). Modulation detection in subjects with relatively flat hearing losses. *Journal of Speech, Language, and Hearing Research*, *35*(3), 642–653.

Banks, W. P., Roberts, D., & Ciranni, M. (1995). Negative priming in auditory attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1354.

Berger, H. (1930). Ueber das elektrenkephalogramm des menschen. *Journal für Psychologie und Neurologie.*

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association*, *82*(397), 112–122.

Beutner, D., Voets, T., Neher, E., & Moser, T. (2001). Calcium dependence of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse. *Neuron*, *29*(3), 681-690. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0896627301002434` doi: https://doi.org/10.1016/S0896-6273(01)00243-4

Bidet-Caulet, A., Mikyska, C., & Knight, R. T. (2010). Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. *NeuroImage*, *50*(1), 277–284.

Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2016). Auditory-inspired speech envelope extraction methods for improved eeg-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(5), 402–412.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707.

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Rabinowitz, W. M., & Reed, C. M. (1979). Hearing aids–a review of past research on linear amplification, amplitude compression, and frequency lowering. *ASHA monographs*(19), 1–114.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433–436.

Bregman, A. S. (1978). The formation of auditory streams. In *Attention and performance vii* (pp. 63–75).

Routledge.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound.* MIT press.

Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics*, *37*(5), 483–493.

Broadbent, D. (1958). Perception and communication.

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*(24), 3976–3983.

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25–31.

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.

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*(3), 1101–1109.

Buchner, A., & Mayr, S. (2004). Auditory negative priming in younger and older adults. *The Quarterly Journal of Experimental Psychology Section A*, *57*(5), 769–787.

Buus, S., & Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *JARO: Journal of the Association for Research in Otolaryngology*, *3*(2), 120.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in cognitive sciences*, *8*(10), 465–471.

Chalas, N., Daube, C., Kluger, D. S., Abbasi, O., Nitsch, R., & Gross, J. (2023). Speech onsets and sustained speech contribute differentially to delta and theta speech tracking in auditory cortex. *Cerebral Cortex*.

Chelazzi, L., Marini, F., Pascucci, D., & Turatto, M. (2019). Getting rid of visual distractors: The why, when, how, and where. *Current opinion in psychology*, *29*, 135–147.

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.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, *118*(2), 887–906.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287–314.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current biology*, *14*(19), R850–R852.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic bulletin & review*, *8*(2), 331–335.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli. *Frontiers in human neuroscience*, *10*, 604.

Dalton, P., & Lavie, N. (2004). Auditory attentional capture: effects of singleton distractor sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 180.

Daly, H. R., & Pitt, M. A. (2021). Distractor probability influences suppression in auditory selective attention. *Cognition*, *216*, 104849.

Darwin, C., & Hukin, R. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *The Journal of the Acoustical Society of America*, *107*(2), 970–977.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892–2905.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *The Journal of the Acoustical Society of America*, *102*(5), 2906–2919.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423–3431.

Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., ... Munro, K. J. (2015). Hearing loss and cognition: the role of hearing aids, social isolation and depression. *PloS one*, *10*(3), e0119616.

De Boer, E., & Kuyper, P. (1968). Triggered correlation. *IEEE Transactions on Biomedical Engineering*(3), 169–179.

Decruy, L., Vanthornhout, J., & Francart, T. (2019). Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *Journal of neurophysiology*, *122*(2), 601–615.

Desimone, R., Duncan, J., et al. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193–222.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological review*, *70*(1), 80.

Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457–2465.

Dillon, H. (1996). Tutorial compression? yes, but for low or high frequencies, for low or high intensities, and with what response times? *Ear and hearing*, *17*(4), 287–307.

Dillon, H. (2008). Hearing aids. Hodder Arnold.

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.

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.

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.

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in human neuroscience*, *8*, 311.

Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word* (Vol. 14). MIT press Cambridge, MA.

Drennan, D. P., & Lalor, E. C. (2019). Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. *eneuro*, *6*(3).

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America*, *97*(1), 585–592.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680.

Duncan, J. (2006). Eps mid-career award 2004: brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology*, *59*(1), 2–27.

Eben, C., Koch, I., Jolicoeur, P., & Nolden, S. (2020). The persisting influence of unattended auditory information: Negative priming in intentional auditory attention switching. *Attention, Perception, & Psychophysics*, *82*(4), 1835–1846.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual review of psychology*, *48*(1), 269–297.

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und forschungsmethoden.*

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*(29), 5750–5759.

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-eeg detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of neural engineering*, *14*(3), 036020.

Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage*, *186*, 33–42.

Forschack, N., Gundlach, C., Hillyard, S., & Müller, M. M. (2022). Electrophysiological evidence for target facilitation without distractor suppression in two-stimulus search displays. *Cerebral Cortex.*

Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, *2*(2), 145–173.

Fox, J. (2015). *Applied regression analysis and generalized linear models.* Sage Publications.

Fox, J. (2016). *Applied regression analysis and generalized linear models.* SAGE Publications.

Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and implications for selective attention. *Psychonomic bulletin & review*, *22*(6), 1577–1597.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience*, *6*(11), 1216–1223.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current opinion in neurobiology*, *17*(4), 437–455.

Fuglsang, S. A., Märcher-Rørsted, J., Dau, T., & Hjortkjær, J. (2020). Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. *Journal of Neuroscience*, *40*(12), 2562–2572.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct evidence for active suppression of salient-but-irrelevant sensory inputs. *Psychological science*, *26*(11), 1740–1750.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2017). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception, & Psychophysics*, *79*(1), 45–62.

Gaspelin, N., & Luck, S. J. (2018). The role of inhibition in avoiding distraction by salient stimuli. *Trends in cognitive sciences*, *22*(1), 79–92.

Gaspelin, N., & Luck, S. J. (2019). Inhibition as a potential resolution to the attentional capture debate. *Current opinion in psychology*, *29*, 12.

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J.-C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing research*, *231*(1-2), 32–41.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of cognitive neuroscience*, *17*(3), 507–517.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge university press.

Geng, J. J. (2014). Attentional mechanisms of distractor suppression. *Current Directions in Psychological Science*, *23*(2), 147–153.

Gevins, A., Leong, H., Smith, M. E., Le, J., & Du, R. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in neurosciences*, *18*(10), 429–436.

Giannoulis, D., Massberg, M., & Reiss, J. D. (2012). Digital dynamic range compressor design—a tutorial and analysis. *Journal of the Audio Engineering Society*, *60*(6), 399–408.

Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Hearing loss is associated with delayed neural responses to continuous speech. *European Journal of Neuroscience*, *55*(6), 1671–1690.

Gillis, M., Kries, J., Vandermosten, M., & Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. *Neuroimage*, *267*, 119841.

Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of neurophysiology*, *84*(3), 1588–1598.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, *15*(4), 511–517.

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

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... others (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980–991.

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2019). The association between hearing impairment and neural envelope encoding at different ages. *Neurobiology of aging*, *74*, 202–212.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, *111*(3), 1340–1348.

Gundlach, C., Forschack, N., & Müller, M. M. (2022). Suppression of unattended features is independent of task relevance. *Cerebral cortex*, *32*(11), 2437–2446.

Hall III, J. W., & Grose, J. H. (1990). Comodulation masking release and auditory grouping. *The Journal of the Acoustical Society of America*, *88*(1), 119–125.

Hall III, J. W., & Grose, J. H. (1994). Signal detection in complex comodulated backgrounds by normal-hearing and cochlear-impaired listeners. *The Journal of the Acoustical Society of America*, *95*(1), 435–443.

Hambrook, D. A., & Tata, M. S. (2019). The effects of distractor set-size on neural tracking of attended speech. *Brain and language*, *190*, 1–9.

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, cognition and neuroscience*, *35*(5), 573–582.

Handy, T. C. (2005). Basic principles of erp quantification. *Event-related potentials: A methods handbook*, 33–55.

Har-shai Yahav, P., & Zion Golumbic, E. (2021). Linguistic processing of task-irrelevant speech at a cocktail party. *Elife*, *10*, e65096.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

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

Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage*, *181*, 617–626.

Hausfeld, L., Shiell, M., Formisano, E., & Riecke, L. (2021). Cortical processing of distracting speech in noisy auditory scenes depends on perceptual demand. *Neuroimage*, *228*, 117670.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393–402.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180.

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Frontiers in systems neuroscience*, *11*, 61.

Holtze, B., Jaeger, M., Debener, S., Adiloğlu, K., & Mirkovic, B. (2021). Are they calling my name? attention capture is reflected in the neural tracking of attended and ignored speech. *Frontiers in neuroscience*, *15*, 643705.

Homans, N. C., Metselaar, R. M., Dingemanse, J. G., van der Schroeff, M. P., Brocaar, M. P., Wieringa, M. H., ... Goedegebure, A. (2017). Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. *The Laryngoscope*, *127*(3), 725–730.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual cognition*, *5*(1-2), 183–216.

Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of neurophysiology*, *109*(12), 3082–3093.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Houghton, G., & Tipper, S. P. (1984). A model of inhibitory mechanisms in selective attention.

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*(5), 2500–2511.

Howard, M. F., & Poeppel, D. (2012). The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. *Neuroimage*, *60*(4), 2118–2127.

Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: perceptual load influences early versus late attentional selection in children and adults. *Developmental psychology*, *38*(3), 363.

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., . . . Wingfield, A. (2012). Central presbycusis: a review and evaluation of the evidence. *Journal of the American Academy of Audiology*, *23*(08), 635–666.

Jaeger, M., Bleichner, M. G., Bauer, A.-K. R., Mirkovic, B., & Debener, S. (2018). Did you listen to the beat? auditory steady-state responses in the human electroencephalogram at 4 and 7 hz modulation rates reflect selective attention. *Brain Topography*, *31*, 811–826.

JASP Team. (2023). *JASP (Version 0.17)[Computer software].* Retrieved from `https://jasp-stats.org/`

Jeffreys, H. (1998). *The theory of probability.* OuP Oxford.

Jensen, N. S., Høydal, E. H., Branda, E., & Weber, J. (2021). Augmenting speech recognition with a new split-processing paradigm. *Hearing Review*, *28*(6), 24–27.

Jessen, S., Obleser, J., & Tune, S. (2021). Neural tracking in infants–an analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience*, *52*, 101034.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological science*, *24*(10), 1995–2004.

Kadir, S., Kaza, C., Weissbart, H., & Reichenbach, T. (2019). Modulation of speech-in-noise comprehension through transcranial current stimulation with the phase-shifted speech envelope. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(1), 23–31.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773–795.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kates, J. M. (2005). Principles of digital dynamic-range compression. *Trends in amplification*, *9*(2), 45–76.

Kates, J. M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *International journal of audiology*, *49*(6), 395–409.

Keefe, J. M., & Störmer, V. S. (2021). Lateralized alpha activity and slow potential shifts over visual cortex track the time course of both endogenous and exogenous orienting of attention. *NeuroImage*, *225*, 117495.

Kemp, D. (1986). Otoacoustic emissions, travelling waves and cochlear mechanisms. *Hearing research*, *22*(1-3), 95–104.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, *30*(2), 620–628.

Kollmeier, B., Peissig, J., & Hohmann, V. (1993). Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *Journal of rehabilitation research and development*, *30*, 82–82.

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.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, *93*(3), 480–490.

Kraus, F., Tune, S., Ruhe, A., Obleser, J., & Wöstmann, M. (2021). Unilateral acoustic degradation delays attentional separation of competing speech. *Trends in Hearing*, *25*, 23312165211013242.

Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, *48*(10), 1217–1232.

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761.

Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European journal of neuroscience*, *31*(1), 189–193.

Lalor, E. C., Power, A. J., Reilly, R. B., & Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of neurophysiology*, *102*(1), 349–359.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Lawrence, B. J., Jayakody, D. M., Bennett, R. J., Eikelboom, R. H., Gasson, N., & Friedland, P. L. (2020). Hearing loss and depression in older adults: a systematic review and meta-analysis. *The Gerontologist*, *60*(3), e137–e154.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., ... others (2020). Dementia prevention, intervention, and care: 2020 report of the lancet commission. *The Lancet*, *396*(10248), 413–446.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010.

Madisetti, V. K. (2018). *The digital signal processing handbook-3 volume set*. CRC press.

Makov, S., Pinto, D., Yahav, P. H.-s., Miller, L. M., & Golumbic, E. Z. (2023). "unattended, distracting or irrelevant": Theoretical implications of terminological choices in auditory selective attention research. *Cognition*, *231*, 105313.

Makov, S., & Zion Golumbic, E. (2020). Irrelevant predictions: distractor rhythmicity modulates neural encoding in auditory cortex. *Cerebral Cortex*, *30*(11), 5792–5805.

Marinato, G., & Baldauf, D. (2019). Object-based attention in complex, naturalistic auditory streams. *Scientific reports*, *9*(1), 2854.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, *164*(1), 177–190.

Marmarelis, V. Z. (2004). *Nonlinear dynamic modeling of physiological systems* (Vol. 10). John Wiley & Sons.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in cognitive sciences*, *9*(6), 296–305.

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological bulletin*, *118*(1), 35.

Mayr, S., & Buchner, A. (2007). Negative priming as a memory phenomenon. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 35–51.

Mayr, S., Buchner, A., Möller, M., & Hauke, R. (2011). Spatial and identity negative priming in audition: Evidence of feature binding in auditory spatial memory. *Attention, Perception, & Psychophysics*, *73*(6), 1710–1732.

Mazza, V., Turatto, M., & Caramazza, A. (2009). Attention selection, distractor suppression and n2pc. *cortex*, *45*(7), 879–890.

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, *19*(1), 431–441.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *The Journal of the Acoustical Society of America*, *79*(3), 702–711.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236.

Micheyl, C., Carlyon, R. P., Shtyrov, Y., Hauk, O., Dodson, T., & Pullvermüller, F. (2003). The neurophysiological basis of the auditory continuity illusion: a mismatch negativity study. *Journal of cognitive neuroscience*, *15*(5), 747–758.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, *14*(2), 247–279.

Moore, B. C. (1990). How much do we gain by gain control in hearing aids? *Acta Oto-Laryngologica*, *109*(sup469), 250–256.

Moore, B. C. (2003). Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. *Speech communication*, *41*(1), 81–91.

Moore, B. C., Glasberg, B. R., & Stone, M. A. (2003). Why are commercials so loud?'perception and modeling of the loudness of amplitude-compressed speech. *Journal of the Audio Engineering Society*, *51*(12), 1123–1132.

Moore, B. C., & Moore, G. A. (2003). Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hearing research*, *182*(1-2), 153–163.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology*, *11*(1), 56–60.

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, *352*, 40–48.

Neill, W. T., Valdes, L. A., Terry, K. M., & Gorfein, D. S. (1992). Persistence of negative priming: Ii. evidence for episodic trace retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 993.

Nelken, I., & Bar-Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Frontiers in neuroscience*, 9.

Noonan, M. P., Crittenden, B. M., Jensen, O., & Stokes, M. G. (2018). Selective inhibition of distracting input. *Behavioural brain research*, *355*, 36–47.

Nourski, K. V., Steinschneider, M., McMurray, B., Kovach, C. K., Oya, H., Kawasaki, H., & Howard III, M. A. (2014). Functional organization of human auditory cortex: investigation of response latencies through direct recordings. *Neuroimage*, *101*, 598–609.

Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Frontiers in human neuroscience*, *6*, 250.

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in cognitive sciences*, *23*(11), 913–926.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., … Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495.

Oliveira, A. J. (1989). A feedforward side-chain limiter/compressor/de-esser with improved flexibility.

Journal of the Audio Engineering Society, *37*(4), 226–240.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1–9.

Organization, W. H., et al. (2017). *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a who report, 2017.* World Health Organization.

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.

Oxenham, A. J., & Bacon, S. P. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear and hearing*, *24*(5), 352–366.

O'Sullivan, J., Chen, Z., Herrero, J., McKhann, G. M., Sheth, S. A., Mehta, A. D., & Mesgarani, N. (2017). Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *Journal of neural engineering*, *14*(5), 056001.

Parthasarathy, A., Bartlett, E. L., & Kujawa, S. G. (2019). Age-related changes in neural coding of envelope cues: peripheral declines and central compensation. *Neuroscience*, *407*, 21–31.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., ... Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS biology*, *10*(1), e1001251.

Patuzzi, R., Yates, G., & Johnstone, B. (1989). Outer hair cell receptor current and sensorineural hearing loss. *Hearing research*, *42*(1), 47–72.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in psychology*, *3*, 320.

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*(6), 1378–1387.

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.

Peelle, J. E., & Wingfield, A. (2016). The neural consequences of age-related hearing loss. *Trends in neurosciences*, *39*(7), 486–497.

Pelli, D. G., & Vision, S. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437–442.

Penn, O. (2020). iboot: Iterated bootstrap for small samples and samples with complex dependence structures. *Journal of Open Source Software*, *5*(50), 2105.

Petersen, E. B. (2022). Hearing-aid directionality improves neural speech tracking in older hearing-impaired listeners. *Trends in Hearing*, *26*, 23312165221099894.

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.

Peterson, A. J., Irvine, D. R., & Heil, P. (2014). A model of synaptic vesicle-pool depletion and replenishment can account for the interspike interval distributions and nonrenewal properties of spontaneous spike trains of auditory-nerve fibers. *Journal of Neuroscience*, *34*(45), 15097–15109.

Pick, G., Evans, E., & Wilson, J. (1977). Frequency resolution in patients with hearing loss of cochlear origin. *Psychophysics and physiology of hearing*, 273–281.

Picton, T., Alain, C., Woods, D. L., John, M., Scherg, M., Valdes-Sosa, P., … Trujillo, N. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology*, *4*(2), 64–79.

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. i: Evaluation of components. *Electroencephalography and clinical neurophysiology*, *36*, 179–190.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Team, R. C., et al. (2007). Linear and nonlinear mixed effects models. *R package version*, *3*(57), 1–89.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America*, *83*(6), 2322–2327.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature reviews neuroscience*, *21*(6), 322–334.

Ponjavic-Conte, K. D., Hambrook, D. A., Pavlovic, S., & Tata, M. S. (2013). Dynamics of distraction: competition among auditory streams modulates gain and disrupts inter-trial phase coherence in the human electroencephalogram. *PloS one*, *8*(1), e53953.

Presacco, A., Simon, J. Z., & Anderson, S. (2016). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *Journal of neurophysiology*, *116*(5), 2346–2355.

Presacco, A., Simon, J. Z., & Anderson, S. (2019). Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PloS one*, *14*(3), e0213899.

Puvvada, K. C., & Simon, J. Z. (2017). Cortical representations of speech in a multitalker auditory scene. *Journal of Neuroscience*, *37*(38), 9189–9196.

R Core Team. (2021). *R: A language and environment for statistical computing.* Computer software. Retrieved from `https://www.R-project.org/`

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). sage.

Rennies, J., Verhey, J. L., Appell, J. E., & Kollmeier, B. (2013). Loudness of complex time-varying sounds? a challenge for current loudness models. In *Proceedings of meetings on acoustics ica2013* (Vol. 19, p. 050189).

Rif, J., Hari, R., Hämäläinen, M. S., & Sams, M. (1991). Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and clinical Neurophysiology*, *79*(6), 464–472.

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373.

Rosen, S., & Fourcin, A. (1986). Frequency selectivity and the perception of speech. *Frequency selectivity in hearing*, *373487*.

Rosowski, J. J. (1994). Outer and middle ears. In *Comparative hearing: mammals* (pp. 172–247). Springer.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, *16*, 225–237.

Schmitt, R., Meyer, M., & Giroud, N. (2022). Better speech-in-noise comprehension is associated with enhanced neural speech tracking in older adults with hearing impairment. *cortex*, *151*, 133–146.

Schneider, D., Herbst, S. K., Klatt, L.-I., & Wöstmann, M. (2022). Target enhancement or distractor suppression? functionally distinct alpha oscillations form the basis of attention. *European Journal of Neuroscience*, *55*(11-12), 3256–3265.

Schreiner, C. E., Read, H. L., & Sutter, M. L. (2000). Modular organization of frequency integration in primary auditory cortex. *Annual review of neuroscience*, *23*(1), 501–529.

Schreiner, C. E., & Urbas, J. V. (1986). Representation of amplitude modulation in the auditory cortex of the cat. i. the anterior auditory field (aaf). *Hearing research*, *21*(3), 227–241.

Schwartz, Z. P., & David, S. V. (2018). Focal suppression of distractor sounds by selective attention in auditory cortex. *Cerebral Cortex*, *28*(1), 323–339.

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *Journal of Neuroscience*, *32*(34), 11812–11819.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in neurosciences*, *34*(3), 114–123.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Shinn-Cunningham, B., Best, V., & Lee, A. K. (2017). Auditory object formation and selection. In *The auditory system at the cocktail party* (pp. 7–40). Springer.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in cognitive sciences*, *12*(5), 182–186.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in amplification*, *12*(4), 283–299.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Shuai, L., & Elhilali, M. (2014). Task-dependent neural representations of salient events in dynamic auditory scenes. *Frontiers in neuroscience*, *8*, 203.

Sievers, E. (1876). *Grundzüge der lautphysiologie, zur einführung in das studium der lautlehre der indogermanischen sprachen* (Vol. 1). Breitkopf und Ha rtel.

Simon, J. Z., Depireux, D. A., Klein, D. J., Fritz, J. B., & Shamma, S. A. (2007). Temporal symmetry in primary auditory cortex: implications for cortical connectivity. *Neural computation*, *19*(3), 583–638.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in psychology*, *3*, 15.

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in amplification*, *6*(4), 131–165.

Stevens, J. C., & Marks, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics*, *27*, 379–389.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The journal of the acoustical society of america*, *8*(3), 185–190.

Stone, M. A., & Moore, B. C. (1992). Syllabic compression: Effective compression ratios for signals modulated at different rates. *British Journal of Audiology*, *26*(6), 351–361.

Stone, M. A., & Moore, B. C. (2004). Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *The Journal of the Acoustical Society of America*, *116*(4), 2311–2323.

Stone, M. A., & Moore, B. C. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America*, *121*(3), 1654–1664.

Sweetow, R. W., & Silverman, J. G. (1994). Speech audiometry. In *Handbook of clinical audiology* (pp. 249–264). Lippincott Williams & Wilkins.

The MathWorks. (2021). *MATLAB.* Computer software. Retrieved from `https://www.mathworks.com/products/matlab.html`

Tibshirani, R. (1993). An introduction to the bootstrap. *Statistical Science*, 54–77.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The quarterly journal of experimental psychology*, *37*(4), 571–590.

Tipper, S. P., Weaver, B., & Houghton, G. (1994). Behavioural goals determine inhibitory mechanisms of selective attention. *The Quarterly Journal of Experimental Psychology Section A*, *47*(4), 809–840.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248.

Tukey, J. W., et al. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.

Tune, S., Alavash, M., Fiedler, L., & Obleser, J. (2021). Neural attentional-filter mechanisms of listening success in middle-aged and older individuals. *Nature communications*, *12*(1), 1–14.

Turner, C., Souza, P., & Forget, L. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *97*(4), 2568–2576.

Uchida, Y., Sugiura, S., Nishita, Y., Saji, N., Sone, M., & Ueda, H. (2019). Age-related hearing loss and cognitive decline—the potential mechanisms linking the two. *Auris Nasus Larynx*, *46*(1), 1–9.

Van Canneyt, J., Wouters, J., & Francart, T. (2021). Cortical compensation for hearing loss, but not age, in neural tracking of the fundamental frequency of the voice. *Journal of Neurophysiology*, *126*(3), 791–802.

van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*(1), 204–221.

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Frontiers in neuroscience*, *13*, 977.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179.

Verhulst, S., Altoe, A., & Vasilkov, V. (2018). Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing research*, *360*, 55–75.

Verschueren, E., Vanthornhout, J., & Francart, T. (2021). The effect of stimulus intensity on neural envelope tracking. *Hearing Research*, *403*, 108175.

Verschuure, J., Maas, A., Stikvoort, E., De Jong, R., Goedegebure, A., & Dreschler, W. (1996). Compression and its effect on the speech signal. *Ear and Hearing*, *17*(2), 162–175.

Villchur, E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *The Journal of the Acoustical Society of America*, *53*(6), 1646–1657.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, *54*(3), 426–482.

Wang, B., & Theeuwes, J. (2018a). How to inhibit a distractor location? statistical learning versus active, top-down suppression. *Attention, Perception, & Psychophysics*, *80*, 860–870.

Wang, B., & Theeuwes, J. (2018b). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13.

Waschke, L., Wöstmann, M., & Obleser, J. (2017). States and traits of neural irregularity in the age-varying human brain. *Scientific reports*, *7*(1), 1–12.

Weinstein, B. E., Sirow, L. W., & Moser, S. (2016). Relating hearing aid use to social and emotional loneliness in older adults. *American Journal of Audiology*, *25*(1), 54–61.

Wong, D. D., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., & De Cheveigne, A. (2018). A comparison of regularization methods in forward and backward models for auditory attention decoding. *Frontiers in neuroscience*, *12*, 531.

Wöstmann, M., Alavash, M., & Obleser, J. (2019). Alpha oscillations in the human brain implement distractor suppression independent of target selection. *Journal of Neuroscience*, *39*(49), 9797–9805.

Wöstmann, M., Fiedler, L., & Obleser, J. (2017). Tracking the signal, cracking the code: Speech and speech comprehension in non-invasive human electrophysiology. *Language, Cognition and Neuroscience*, *32*(7), 855–869.

Wöstmann, M., Störmer, V. S., Obleser, J., Andersen, S., Gaspelin, N., Geng, J., . . . others (2022). Ten simple rules to study distractor suppression. *Progress in neurobiology*, 102269.

Wöstmann, M., Waschke, L., & Obleser, J. (2019). Prestimulus neural alpha power predicts confidence in discriminating identical auditory stimuli. *European Journal of Neuroscience*, *49*(1), 94–105.

Zou, J., Xu, C., Luo, C., Jin, P., Gao, J., Li, J., . . . Luo, B. (2021). θ-band cortical tracking of the speech envelope shows the linear phase property. *Eneuro*, *8*(4).

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. *Psychological review*, *72*(1), 3.

Table of figures

1 Measurement of the impule resonse of a LTI System . . . 23
2 Encoding decoding approach . . . 24
3 Dynamic range compressor . . . 28
4 Human auditory periphery model Verhulst . . . 30
5 Bayes factor example . . . 34
6 Experimental design and hypothetical results . . . 39
7 Behavioural results and ERPs to repeats . . . 46
8 TRF Topo . . . 48
9 Neural tracking . . . 49
10 Brain–behaviour relation . . . 51
11 TRFs across trial duration . . . 53
12 Audiogram and PTA . . . 63
13 Stimulus processing pipeline . . . 65
14 Experimental paradigm . . . 67
15 Pilot study . . . 71
16 NH: Behavioural results . . . 74
17 NH:Neural response . . . 76
18 NH:Neural response pairs . . . 77
19 NH: Comparison attentional pairs . . . 79
20 Verhulst modeling: AN and EFR . . . 81
21 HI: Behavioural results . . . 83
22 HI: Neural response . . . 84
23 HI: Neural response pairs . . . 85
24 HI: Comparison attentional pairs . . . 87
25 Neural response HI vs. NH . . . 88
26 Example loudness matching . . . 91
27 Loudness ratings . . . 92
28 Loudness as function of conditions . . . 93

Summary

People are often confronted with complex auditory environments where multiple sound sources compete for their attention. The ability to selectively attend to one particular sound source over others is essential for effective communication and even mild to moderate hearing loss can impair the processing of speech in such environments. Cherry (1953) introduced the concept of the cocktail party problem, and since then, research has focused on understanding the neural mechanisms underlying selective attention. Auditory object formation is a crucial process in the perception of continuous speech and the selective attention required to solve the cocktail party problem (Bregman, 1994; B. Shinn-Cunningham et al., 2017; Bizley & Cohen, 2013). Selective attention refers to the ability to control which information to attend to in the presence of distractors (Desimone et al., 1995). The mechanism of how selective attention is implemented is a topic of ongoing debate in attention research, and it is proposed that a pre-defined baseline is required to distinguish target enhancement and distractor suppression (Wöstmann et al., 2022). The present work investigated these sub-processes of selective attention in young and normal hearing participants using a psychophysically augmented continuous speech paradigm. The experiment involved using a speech stimulus that was irrelevant and served as a baseline, against which the processing of relevant target speech and irrelevant distractor speech could be compared. Participants had to continuously monitor and detect repeated segments in the relevant speech while ignoring repeats in the irrelevant speech. This helped to determine if the neural responses to relevant, irrelevant, and baseline speech could explain the variation in attentional performance on a trial-by-trial basis.

Hearing-impaired individuals often struggle in multi-talker environments, making them an interesting test case for investigating the neural dynamics of selective attention in complex listening scenarios. Hearing aids are the most common treatment for individuals with presbycusis, and new hearing aid technology enables independent processing of sound from different directions. Studies have shown that the temporal speech envelope is crucial for speech comprehension (Shannon et al., 1995; Peelle et al., 2010; Ding & Simon, 2014). Our hypothesis is that dynamic range compression, a signal processing technique that directly affects the temporal speech envelope, impairs both behavioural performance and neural speech tracking. Additionally, we propose that applying compression only to ignored talkers in a multi-talker situation will increase the neural separation between the attended and ignored talkers and lead to improved behavioural performance. We employed an adapted version of the psychophysically augmented continuous speech paradigm. In contrast to the first study, we reduced the number of speakers and modified the positions of the two competing speakers to the front and back, mimicking the processing of hearing aids. The compression was applied randomly and equally in advance to the speech streams.

In Study 1, it was discovered that selective attention is achieved by enhancing the target speech, rather than suppressing the distraction. The study involved 19 young adults. All participants reported having German as their native language and having normal hearing. The study's results revealed that listeners made more false alarms from the distractor speech than from the neutral stream. However, the neural

representation of target speech was strengthened, and no suppression of distraction was observed below the neutral baseline. Only, neural tracking to target speech explains performance. These findings imply that the primary mechanism behind selective attention is the enhancement of the target.

The second part of the thesis conducted several studies to investigate the effects of amplitude compression and selective attention on neural separation and behavioural response in normal hearing and hearing-impaired participants. In Study 2, a pilot study (N=6) was conducted with different compression and expansion ratios to determine an appropriate ratio to be used in the follow-up study. The findings showed that a 1:8 compression ratio significantly reduced the brain's ability to track speech. Study 3 used a continuous speech paradigm and revealed that compression on both attended and ignored streams decreased behavioural performance and neural speech tracking while increasing neural separation when only the ignored stream was compressed in N = 24 normal hearing participants. In Study 4, a computational model of the human auditory periphery was used to simulate the firing rate of the auditory nerve and the envelope following response for both normal hearing and hearing-impaired participants. The simulation results indicated that changes in the auditory periphery did not confound the effects observed in the neural speech tracking study and the follow-up study for hearing-impaired participants. Study 5 aimed to study the impact of amplitude compression and selective attention on neural separation and behavioural response in individuals with hearing impairment (N=7). The results showed comparable patterns to those observed in normal-hearing participants, with reduced performance and neural tracking for amplitude-compressed speech. However, unlike the normal-hearing participants, the results showed enhanced neural speech tracking for the attended stream when only the ignored stream was compressed.

This thesis investigates the sub-processes of selective attention using speech streams utilising augmented psychophysically speech paradigms. The first part examines top-down selective attention mechanisms and finds evidence for target enhancement but not for distractor suppression. The second part uses dynamic range compression as an acoustic manipulation and shows some indications of suppressive mechanisms in neural speech tracking. The absence of evidence for distractor suppression in the first part may be due to different neural mechanisms implementing it. Proactive suppression is associated with alpha power, while reactive suppression is more reflected in the neural tracking of speech. Reactive suppression was found to be absent in auditory cortex responses in a multi-talker situation. Future studies are needed to investigate proactive suppression using other neural measures. The second part demonstrates that dynamic range compression on ignored speech could increase neural tracking of attended speech, suggesting a possible suppression mechanism. As in the first part, a baseline control condition is needed to reliably distinguish target enhancement from distractor suppression. These findings indicate the complexity of selective attention and the need for further research to understand its neural mechanisms. Researchers could also investigate the effect of compression on other behavioural measures, explore the impact of modulation coherence on selective attention, and consider the potential benefits of side chain compression in hearing aids.

Zusammenfassung

Menschen werden oft mit komplexen akustischen Umgebungen konfrontiert, in denen mehrere Schallquellen um ihre Aufmerksamkeit konkurrieren. Die Fähigkeit, sich selektiv auf eine bestimmte Schallquelle zu konzentrieren und andere auszublenden, ist für eine effektive Kommunikation unerlässlich. Selbst geringgradige Hörverluste können die Verarbeitung von Sprache in solchen Umgebungen beeinträchtigen. Cherry (1953) führte das Konzept des Cocktailparty-Problems ein, und seither konzentriert sich die Forschung darauf, die neuronalen Mechanismen zu verstehen, die der selektiven Aufmerksamkeit zugrunde liegen. Die Bildung auditiver Objekte ist ein entscheidender Prozess bei der Wahrnehmung kontinuierlicher Sprache und der selektiven Aufmerksamkeit, die für die Lösung des Cocktailparty-Problems erforderlich ist (Bregman, 1994; B. Shinn-Cunningham et al., 2017; Bizley & Cohen, 2013). Selektive Aufmerksamkeit bezieht sich auf die Fähigkeit, zu kontrollieren, auf welche Informationen man sich in Gegenwart von Störgeräuschen konzentriert (Desimone et al., 1995). Der Mechanismus, wie selektive Aufmerksamkeit implementiert wird, ist ein Thema der laufenden Debatte in der Aufmerksamkeitsforschung, und es wird vorgeschlagen, dass eine vorgegebene Baseline erforderlich ist, um Zielverstärkung und Distraktorsuppression zu unterscheiden (Wöstmann et al., 2022). Die vorliegende Arbeit untersuchte diese Teilprozesse der selektiven Aufmerksamkeit bei jungen und normal hörenden Teilnehmern unter Verwendung eines psychophysisch erweiterten kontinuierlichen Sprachparadigmas. Das Experiment beinhaltete die Verwendung eines Sprachstimulus, der irrelevant war und als Baseline diente, um die Verarbeitung relevanter Zielsprache und irrelevanter Distraktorsprache vergleichen zu können. Die Teilnehmer mussten fortlaufend wiederholte Abschnitte in der relevanten Sprache erkennen und gleichzeitig Wiederholungen in der irrelevanten Sprache ignorieren. Dies half dabei zu bestimmen, ob die neuronalen Antworten auf relevante, irrelevante und Baseline-Sprache die Varianz in der Aufmerksamkeitsleistung auf einer Trial-by-Trial-Basis erklären konnten.

Hörgeschädigte Personen haben oft Schwierigkeiten in Situationen mit mehreren Sprechern, was sie zu einem interessanten Testfall für die Untersuchung der neuronalen Dynamik der selektiven Aufmerksamkeit in komplexen Hörszenarien macht. Hörgeräte sind die häufigste Behandlung für Menschen mit Presbyakusis, und neue Hörgerätetechnologien ermöglichen die unabhängige Verarbeitung von Schall aus verschiedenen Richtungen. Studien haben gezeigt, dass die zeitliche Sprachhülle für das Sprachverständnis eine wichtige Rolle spielt (Shannon et al., 1995; Peelle et al., 2010; Ding & Simon, 2014). Unsere Hypothese ist, dass die dynamische Bereichskompression, eine Signalverarbeitungstechnik, die direkt die zeitliche Sprachhülle beeinflusst, sowohl die Performance als auch das neuronale Sprachtracking beeinträchtigt. Darüber hinaus schlagen wir vor, dass die Anwendung von Kompression nur auf ignorierte Sprecher in einer Situation mit mehreren Sprechern die neuronale Trennung zwischen dem aufmerksamen und ignorierten Sprecher erhöht und zu einer verbesserten Performance führt. Wir verwendeten eine angepasste Version des psychophysischen augmentierten kontinuierlichen Sprachparadigmas. Im Gegensatz zur ersten Studie haben wir die Anzahl der Sprecher reduziert und die Positionen der beiden konkurrierenden Sprecher vorne und hinten modifiziert, um die Verarbeitung von Hörgeräten zu simulieren. Die Kom-

pressionsmanipulation wurde randomisiert im Voraus auf die Sprachströme angewendet.

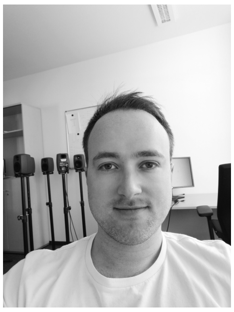
In Studie 1 wurde gezeigt, dass selektive Aufmerksamkeit durch die Verstärkung des Zielsprechers erreicht wird, anstatt den störenden Sprecher zu unterdrücken. Die Studie umfasste 19 junge Erwachsene. Alle Teilnehmer gaben an, Deutsch als Muttersprache zu haben und normales Hörvermögen zu besitzen. Die Ergebnisse der Studie zeigten, dass die Zuhörer mehr False-alarms von dem störenden Sprecher als vom neutralen Sprecher machten. Die neuronale Repräsentation des Zielsprechers wurde verstärkt, und es wurde keine Unterdrückung des störenden Sprechers unterhalb der neutralen Baseline beobachtet. Nur das neuronale Tracking des Zielsprechers erklärte die Performance im Verhalten. Diese Ergebnisse legen nahe, dass der primäre Mechanismus hinter selektiver Aufmerksamkeit die Verstärkung des Zielsprechers ist.

Der zweite Teil der Arbeit führte mehrere Studien durch, um die Auswirkungen der Amplitudenkompression und selektiven Aufmerksamkeit auf die neuronale Trennung und das Verhaltensverhalten bei normal hörenden und hörgeschädigten Teilnehmern zu untersuchen. In Studie 2 wurde eine Pilotstudie (N=6) mit unterschiedlichen Kompressions- und Expansionsverhältnissen durchgeführt, um ein geeignetes Verhältnis für die Folgestudie zu ermitteln. Die Ergebnisse zeigten, dass ein Kompressionsverhältnis von 1:8 die Fähigkeit des Gehirns, Sprache zu verfolgen, signifikant verringerte. Studie 3 verwendete ein kontinuierliches Sprachparadigma und zeigte, dass die Kompression sowohl auf den beachteten als auch auf den ignorierten Streams die Performance und das neuronale Sprachtracking verringerte, während die neuronale Trennung erhöht wurde, wenn nur der ignorierte Stream bei N = 24 normal hörenden Teilnehmern komprimiert wurde. In Studie 4 wurde ein computerbasiertes Modell des menschlichen Hörorgans verwendet, um die Feuerrate des Hörnervs und die Envelope-following-Response für sowohl normal hörende als auch hörgeschädigte Teilnehmer zu simulieren. Die Simulationsergebnisse zeigten, dass Veränderungen im Hörorgan die Auswirkungen der beobachteten Studien zum neuronalen Sprachtracking und Folgestudie für hörgeschädigte Teilnehmer nicht beeinträchtigten. Schließlich zielt Studie 5 darauf ab, die Auswirkungen der Amplitudenkompression und selektiven Aufmerksamkeit auf die neuronale Trennung und das Verhaltensverhalten bei Personen mit Hörbeeinträchtigung (N=7) zu untersuchen. Die Ergebnisse zeigten vergleichbare Muster wie bei normal hörenden Teilnehmern, mit reduzierter Performance und neuronalem Tracking für amplitudenkomprimierte Sprache. Im Gegensatz zu den normal hörenden Teilnehmern zeigten die Ergebnisse jedoch ein verbessertes neuronales Sprachtracking für den beachteten Stream, wenn nur der ignorierte Stream komprimiert wurde.

Diese Arbeit untersucht die Teilprozesse der selektiven Aufmerksamkeit unter Verwendung von psychophysischen Sprachparadigmen. Der erste Teil untersucht top-down selektive Aufmerksamkeitsmechanismen und findet Hinweise auf eine Zielverstärkung, aber nicht auf eine Distraktorsuppression. Der zweite Teil verwendet dynamische Bereichskompression als akustische Manipulation und zeigt einige Anzeichen für supressive Mechanismen im neuronalen Sprachtracking. Das Fehlen von Hinweisen auf Distraktorsuppression im ersten Teil könnte auf verschiedene neuronale Mechanismen zurückzuführen sein, die dies implementieren. Proaktive Suppression ist mit Alpha-Power assoziiert, während reaktive Suppression eher im neuronalen Tracking von Sprache reflektiert wird. Reaktive Suppression im neuronalen Tracking wurde hier in einer Mehrsprecher-Situation nicht gefunden. Zukünftige Studien sind notwendig, um proaktive Suppression mit anderen neuronalen Maßen zu untersuchen. Der zweite Teil zeigt, dass die dynamische Bereichskompression bei ignorierten Sprachsignalen das neuronale Tracking von beachteter Sprache erhöhen kann, was auf einen möglichen Suppressionsmechanismus hinweist. Wie im ersten Teil ist allerdings auch hier eine Kontrollbedingung notwendig, um Zielverstärkung und Distraktorsuppression zuverlässig zu unterscheiden. Diese Ergebnisse zeigen die Komplexität selektiver Aufmerksamkeit und die Notwendigkeit weiterer Forschung, um ihre neuronalen Mechanismen zu verstehen. Zukünftige Studien könnten den Effekt der Kompression auf andere Verhaltensmaße untersuchen, den Einfluss von Comodulation auf selektive Aufmerksamkeit erforschen und die potenziellen Vorteile von Side-chain-Kompression bei Hörgeräten berücksichtigen.

Curriculum vitae

Martin Orf
M. Sc. Auditory Technology



Education and professional experience

since 02/2023	**Post-Doc** at the University of Lübeck, Institute for Psychology
11/2019 - 01/2023	**PhD candidate** at the University of Lübeck, Institute for Psychology, funds by Widex Sivantos Audiology
10/2018 - 11/2019	**Intern, Master's candidate, Master's graduate** at the University of Lübeck, Institute for Psychology
10/2017 - 10/2019	**M. Sc. Auditory Technology,** University of Lübeck, final grade: excellent (A)
10/2014 - 10/2017	**B. Sc. Hearing Acoustics,** Technical University of Lübeck, final grade: excellent (A)
10/2011 - 10/2014	**Apprentice** at vocational school for hearing aid acoustics and at the hearing aid company: "Augenoptik und Hörakustik Maaß", Bad Hersfeld, Germany
10/2008 - 10/2011	**Modellschule Obersberg, high school,** Bad Hersfeld, Germany

Publication

Orf, M., Wöstmann, M., Hannemann, R., Obleser, J., (in review) The cortical neural tracking response reflects target enhancement but not distractor suppression in a psychophysically augmented continuous-speech paradigm. *iScience*

Talks and poster

Talks

- Audiological Research Unit, Erlangen (virtual), 2020
- ISAAR International Symposium on Auditory and Audiological Research, Nyborg (virtual), 2021
- Audiological Research Unit, Erlangen (virtual), 2021
- Audiological Research Unit, Erlangen (virtual), 2022
- DGA, Erfurt, 2022

Poster

- Speech in Noise, (virtual), 2022

- Annual meeting of Society for Neuroscience 2022 in San Francisco (U.S.)
- Auditory Cortex, Magdeburg, 2022

Teaching

Lectures

Einführung in die Hörakustik (B.Sc. Pflege), 2022, University of Lübeck

Supervision

Supervised master intern, Iris Borschke, in auditory technology