UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Psychologie I

der Universität zu Lübeck

Direktor: Prof. Dr. Nico Bunzeck

Neural signatures of auditory selective attention under dynamic listening conditions

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

vorgelegt von
Lorenz Fiedler
aus Jena
Lübeck, 2018

1. Berichterstatter: Prof. Jonas Obleser

2. Berichterstatter: Prof. Stefan Debener

Tag der mündlichen Prüfung: 24. April 2019

Zum Druck genehmigt. Lübeck, den 25. April 2019

Neural signatures of auditory selective attention under dynamic listening conditions

Acknowledgements

I would like to thank all the people that were involved in the creation of this thesis.

First and foremost, I would like to thank my supervisor Prof. Jonas Obleser, who encouraged and taught an electrical engineer to dive into auditory neuroscience. With his extensive expertise, he guided my work with keen feedback during countless fruitful discussions. Importantly, he has always been open for the unfolding of new ideas that allowed for such an interdisciplinary venture.

A big thank you goes out to all people who have been associated with the *Research Group Auditory Cognition* at the *Max Planck Institute in Leipzig* and the *University of Lübeck.* The animated discussions during group meetings and the spontaneous pow-wows have always been inspiring and helpful. Thank you Malte Wöstmann, Dunja Kunke, Steven Kalinke, Alex Brandmeyer, Sung-Joo Lim, Sophie Herbst, Mohsen Alavash, Leonhard Waschke, Franziska Scharata, Tanja Kruse, Christa Marx, Anne Hermann, Lea Maria Schmitt, Michael Plöchl, Sarah Tune, Jens Kreitewolf, Julia Erb & Felix Deilmann. I am grateful to all the students involved in recording and analyzing hard drives full of data: Sarah Sentis, Simon Grosnick, Cécilia Souyris, Thomas Cuntz, Stephan Müller, Daniel Bank, Raphaela Wurzer and all the other students spending hours in the lab to keep it all up and running.

I would also like to thank all people involved in the collaboration with *Eriksholm Research Centre,* Thomas Lunner, Carina Graversen and Eline Borch Petersen, who provided a welcoming environment for me to record in-ear EEG data.

This thesis would not exist without people outside the lab. I would like to deeply thank Anna Hubner for her infinite support. Beyond his scientific work, I will always very much appreciate the presence of Leonhard Waschke, who has become one of my best friends. I very much enjoyed coming home every night to my flatmates and friends Julia Greiner, Fabian Quiering, Rahel Wacker, Till Ehrmann, Nils Abke and Clara Haug. I would also like to thank my family. Thank you Luise Fiedler to be a source of inspiration for graphic design and for making the first and only in-ear EEG movie. The quality of this manuscript has strongly profited from the feedback of Aviv Hilbig-Boaker.

Lorenz Fiedler, Lübeck, April 26, 2019

Abstract

In natural environments, multiple objects compete for our attention. Thus, the incoming information must be reduced by attentional filtering. Listening to a talker of choice (*top-down* attention) can be corrupted by the attentional capture of a more salient, distracting talker (*bottom-up* attention). In this thesis, I investigated the neural signatures of auditory attention under continuously varying acoustic conditions.

Two neural signatures of auditory attention have been recently studied: the *neural tracking* of speech and the modulation of *alpha power*. *Neural tracking* refers to the neural phase-locking to the (spectro-) temporal fluctuations of speech, which has been previously shown to be indicative of *top-down* attention. *Alpha power* refers to induced neural oscillations around 10 Hz, which has been previously proposed to control the distribution of neural resources through inhibition of brain regions or neural pathways processing irrelevant information. A comprehensive understanding of the interplay between *neural tracking* of speech and the modulation of *alpha power* has not been established yet.

In this thesis I used electroencephalography to investigate the simultaneous attentional modulation of *neural tracking* and *alpha power* during continuous selective listening tasks. We dynamically manipulated the demand on *top-down* attentional control by varying the signal-to-noise ratio (SNR) and the location of the to-be-attended as well as the to-be-ignored talker. We applied and refined *forward encoding models*, which allowed us to predict the neural response to continuous speech as well as to detect the listener's attentional focus. Based on those predicted neural responses, we traced the cortical representation (i.e., *neural tracking*) of the attended and the ignored talker, respectively. We disentangled the impact of *bottom-up* versus *top-down*-attentional modulation (i.e., *neural selectivity*). Furthermore, we investigated whether the neural signatures of auditory attention can be recorded at a reduced set of EEG electrodes, which could provide valuable information for the neural steering of a hearing aid.

In the first part (studies 1–3), I show that the *neural tracking* of speech is composed of *bottom-up*-driven as well as *top-down*-controlled modulation. We show that *top-down* attention strongly shapes later components of the neural response. This results in increased *neural selectivity* by way of suppressed responses to the ignored talker. Under most adverse conditions, increased late *neural tracking* and *neural selectivity* of the ignored talker indicates *top-down*-controlled suppression. However, the

modulation of *alpha power* did not follow the hypothesized direction. Neither the SNR nor the location of the talkers predicted the attentional modulation of *alpha power*.

In the second part of this thesis (studies 4–6), I show that a listener's focus of attention can be detected with in-ear EEG based on the *neural tracking* of speech. We replicate this finding and additionally show that the increased late tracking of the ignored talker is also indicated by in-ear EEG.

I argue here that the *top-down neural tracking* of speech and the modulation of *alpha power* are two distinct neural strategies rather than two sides of the same coin. I conclude that attentional filtering is primarily achieved by spectro-temporal, proactive filtering in auditory cortex and that adverse listening conditions is controlled by an additional late suppressive tracking of the ignored talker in the fronto-parietal attention network. I conclude that neural strategies related to the attentional modulation of *alpha power* were not obligatory under the given task demands. Regarding neurally steered hearing aids, *neural tracking* provides valuable information for the neural steering of a hearing aid.

In sum, the *neural tracking* of speech is the most prominent signature of *top-down* auditory attention and I showed that it adapts to the current listening conditions. I could not confirm a significant role of *alpha power* modulation as a signature of *top-down* attention to continuous speech. It is up to further studies to close the gap between the attentional modulation of the *neural tracking* of speech and of *alpha power* during continuous listening.

Contents

1 General introduction

1.1 Auditory selective attention

In natural environments, multiple objects compete for the allocation of our attention. In a crowded *auditory scene*, the ability to attend to a certain sound source in the presence of multiple distractors has been called the *cocktail-party effect* (Cherry, 1953). During the investigation of selective attention in the last 70 years or so, dichotomous concepts such as *parsing* and *grouping*, *early* and *late* selection, *bottom-up* and *top-down* were coined. Those terms will be described within this section.

1.1.1 Selective attention in an auditory scene

Auditory objects such as a certain talker play an important role when studying auditory selective attention. The process of the perceptual formation of auditory objects was called *auditory scene analysis* by Bregman (1990) and opened a new field of research. It was argued that attention operates on an object-based level (e.g., Duncan, 2006; Shinn-Cunningham, 2008) and it was shown that attention is not a prerequisite for the formation of objects (Alain et al., 2001; Dyson et al., 2005; Hautus and Johnson, 2005). However, it was also shown that the formation of auditory objects depends on attention (Zobel et al., 2015).

The main questions asked in the field of *auditory scene analysis* concerned the features (such as pitch or location) of an incoming sound or sound mixture that may or may not lead to the formation of auditory objects, also called auditory streams. Two conceptual processing stages were introduced to describe the process of auditory scene analysis: First, *parsing* describes the decomposition of a sound or sound mixture into its basic building blocks. *Parsing* starts at the level of the cochlea. The cochlea can be approximated by a bank of bandpass filters which returns a time-frequency representation of the auditory input, which can be illustrated in the form of a spectrogram, also called cochleogram (e.g., Patterson, 1976; Glasberg and Moore, 1990). Each time-frequency bin of such a cochleogram might or might not belong to a certain auditory object. Subsequently, *grouping* reorganizes those bins and ascribes each bin to an auditory stream. In the presence of multiple sound sources, *parsing* is mathematically ill-posed due to the superposition of multiple waveforms. Hence, the auditory system must rely on certain cues that allow

assumptions about the present auditory objects in the *auditory scene* (e.g., Griffiths and Warren, 2004; Bizley and Cohen, 2013).

The features that allow the formation of auditory streams have been categorized by Bregman (1990) into horizontal and vertical cues, which he based on the common illustration of spectrograms. While horizontal cues refer to the temporal regularities of an auditory object, vertical cues refer to its spectral regularities. For example, an alternating sequence of tones of separate pitch might be perceived as two separate streams if the pitch difference exceeds a certain threshold and as one stream otherwise (Micheyl et al., 2005). Interestingly, close to threshold, bi-stability of the perception allows a listener to actively bias the perception, which raised the question if attention can act prior to object formation or only on previously formed objects (Gutschalk et al., 2005; see Snyder et al., 2012 for a review). However, if the tones are temporarily aligned, they unavoidably form one stream in form of a complex tone, which highlights the role of temporal coherence in auditory object formation (Elhilali et al., 2009; Teki et al., 2013; Shamma et al., 2013).

This thesis primarily focuses on the attention-related neural processing of talkers without spatial segregation. Irrespective of the question which cue dominates object formation in this case, we mainly investigate the neural response to the broad-band temporal dynamics of the talkers represented by the envelope of their speech signals.

1.1.2 Early and late selection

Selective attention is a cognitive process which determines how and if a certain stimulus, amongst the presence of others, is transferred towards perception. One of the first models of selective attention was predominantly tested in the auditory modality (Broadbent, 1958). Broadbent presented two separate auditory inputs at each ear (i.e., dichotic listening), while subjects had to attend and later repeat one of the two. Subjects were only able to report the content of the attended input. Broadbent argued that the processing of the incoming information is constrained by limited computational resources and that a filter is tuned to reduce the amount of incoming information. This reduction by way of a filter was later called *bottleneck*. The filter was suggested to act at an early stage (i.e., early selection), which means that it is tuned to basic physical properties (e.g. location or pitch) of a stimulus rather than the content (meaning of the message) it is about to transfer.

The concept of mere early selection was challenged by later findings. Moray (1959) found that subjects are still able to detect their own name in the unattended stream during dichotic listening. Treismann (1960) showed that a sudden switch of the two streams tempts subjects to follow the stream in the to-be-ignored ear, at least for a few words. She argued that the selective filter tuned to physical properties of the stimulus such as its location is not unalterable but can be overruled by semantic cues. Going beyond the concept of an attentional filter that is strictly tuned to the physical properties of a stimulus, the concept of late selection was established as well. The model of Deutsch and Deutsch (1963) even assumed that all input is processed up to the semantic representation and selection merely acts at this stage.

Even if early and late selection are well established terms, it is not well defined where early selection ends and where late selection begins. In sum, the two concepts highlight that the investigation of (auditory) attention always comes with the assumption that the underlying neural processes are organized in a hierarchical structure and that attention transforms the representation of sensory input in an incremental manner. Until now this view has not changed much, but how and in which direction information flows through this hierarchy as well as which stages are involved is still a matter of debate.

1.1.3 Bottom-up and top-down attention

The bi-directional concept of *bottom-up* and *top-down* attention accounts for the fact that in selective attention tasks, to-be-ignored or distracting stimuli can capture attention if those stimuli exceed a certain saliency threshold (for review see Katsuki and Constantinidis, 2014; Wolfe and Horowitz, 2004). While *bottom-up* attention is the exogenous, sensory-driven capture of attention by a stimulus due to its saliency, *top-down* attention describes the internal, voluntary selection of a certain stimulus. The concept of *bottom-up* and *top-down* attention is closely linked to early and late selection (see section 1.1.2), because the features of an intentionally ignored stimulus that could possibly capture *bottom-up* attention are determined by how late (i.e., up to which stage) an ignored stimulus is processed.

Many studies have investigated the interplay between *bottom-up* and *top-down* attention in the visual modality (for review: Egeth and Yantis, 1997, Connor et al., 2004). However, visual attention differs from auditory attention in some substantial respects. For example, the involvement of eye-movements into visual *bottom-up* attention has been extensively studied, but

an equivalent movement of the peripheral sensory organs cannot be found in the auditory modality. Furthermore, dichotic listening tasks have been used multiple times to study auditory attention, whereas presenting two different stimuli to each eye is rather unusual in studies of visual attention. Another aspect is the difference of the temporal structure between visual and auditory stimuli. While visual stimuli are usually presented in a static fashion, auditory stimuli such as speech unfold in time. Consequently, auditory *bottom-up* and *top-down* attention have to be investigated with respect to the unique properties of the to be encoded sensory signals.

Which features of unattended or distracting stimuli attract attention and which intensity of those features are needed to rule out *top-down* attention? To answer this question, researchers tried to come up with objective measures of saliency (for review see Kaya and Elhilali, 2017). One approach is to psycho-acoustically quantify the saliency of a stimulus based on ratings of subjects (Kayser et al., 2005; Tsuchida and Cottrell, 2012; Kaya and Elhilali, 2014). Basic features such as sound intensity, spectral and temporal contrast were shown to strongly contribute to enhanced saliency. However, such designs do not allow to directly infer on *bottom-up* attention, since subjects performed the main task on the rating of an attended stimulus. To overcome this problem, other studies investigated the distractive potential of certain stimulus features such as regularity (Southwell et al. 2017) or verbal and spatial deviance (Vachon et al., 2017). It turned out that the multidimensional space of acoustic features and their probable non-linear interaction hampers the development of an overarching model. In a recent attempt, a comprehensive dataset of auditory scenes was presented in random pairs (one to each ear). Subjects had to freely listen to the scenes and continuously report which of the scenes attracts more of their attention (Huang and Elhilali, 2017). As shown before, change of loudness (i.e., perceived sound intensity) was found to be mostly driving the salience, but also other features such as change of harmonicity, pitch, timbre (i.e., brightness) and spectral frequency (i.e., scale) were found to contribute to overall saliency. This highlights that in particular the rate of change of a feature and not the absolute intensity is crucial for capturing *bottom-up* attention.

Furthermore, the investigation of attention-capturing features led to the conclusion that not only the instantaneous features but also the stimulus history plays an important role, not least because the stimulus history influences predictability of upcoming changes (Kaya and Elhilali, 2017).

1.2 Electroencephalography

Electroencephalography (EEG) is an electrophysiological method of measuring the cortical (and also subcortical) activity of the (human) brain. Technically described, EEG is the voltage fluctuation between at least two electrodes attached to the scalp. The first EEG measurements were conducted by Hans Berger (1929). Since then, the basic principle of EEG has not changed very much, whereas the certainty about the neural source of the EEG signals has increased. Currently it is accepted that EEG mainly captures the shifting of extracellular charge due to the superposition of the postsynaptic potentials of multiple neurons within the cerebral cortex (Zschocke,2012). EEG mainly captures the summed potentials of co-aligned ensembles of pyramidal cells lying in radial orientation relative to the skull. However, tangential sources and deeper sources (e.g., brainstem) are captured as well.

Berger (1929) discovered dominant waves of frequencies around 10 Hz, which he later called alpha-waves (α-waves; Berger, 1932). He found the amplitude of those alpha waves was increased when the eyes of the subjects were closed as compared to opened. In the following years, various phenomena discovered by other researchers led to the subdivision of the EEG frequency band, which led to established terms (whereas the exact frequency range varies between authors): delta-band (δ; 1–3.5 Hz), theta-band (θ; 4–7.5 Hz), alpha-band (α; 8–13 Hz), beta-band (β; 14–30 Hz) and gamma-band (γ; 30–90 Hz; Lopes da Silva, 2013).

Within this thesis, the analysis primarily focuses on lower frequency bands (δ & θ), because they concur with dominant frequencies of speech signals' broad-band amplitude modulation (see section 1.3; Ding et al., 2017) as exemplary shown in Figure 1-1. However, it was also shown that power in the α-band co-varies with attention-dependent factors during listening tasks (see section 1.5). Hence, modulations in the α-band are also under investigation in this thesis.

EEG signals can be analyzed within different domains (i.e., representations), depending on the experimental design and the underlying research question. Two important representations are the time and frequency domain (Wöstmann et al., 2017a).

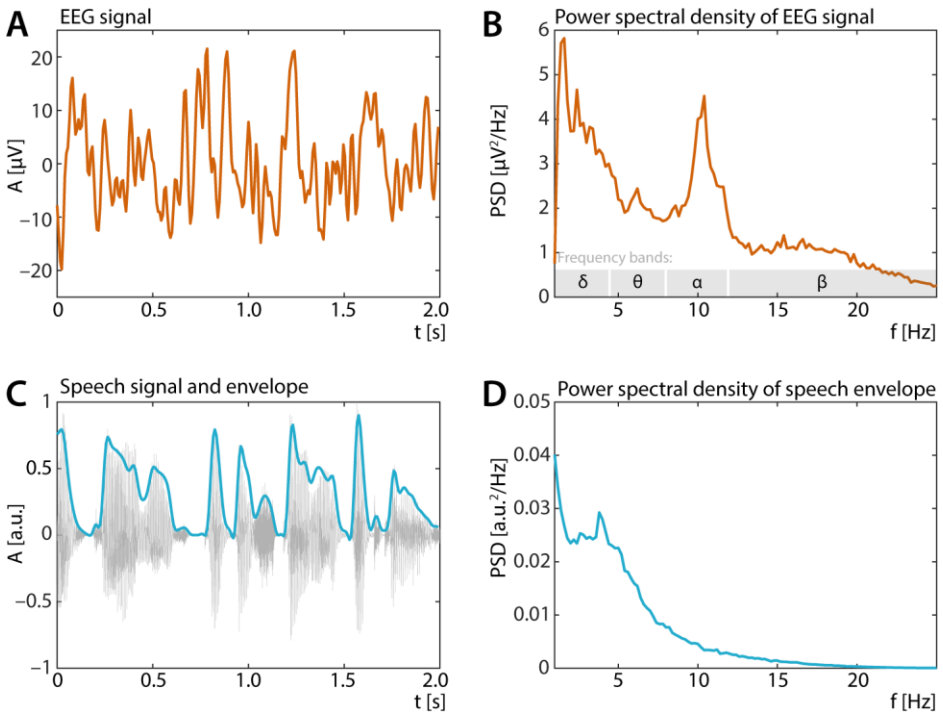

Figure 1-1: Common frequency bands of speech envelope and EEG signal. A) Exemplary EEG signal bandpass-filtered between 1 and 40 Hz. B) Power spectral density of a 15-minute EEG recording. Besides the dominant peak in the α-band, lower frequency bands are showing increased power compared to higher bands. A) The temporal envelope approximates the amplitude modulation of a broad band speech signal. B) The power spectral density of the speech envelope exemplarily shows that amplitude modulation is dominated by frequencies in the range below 10 Hz, which coincides with the lower bands (δ and θ) of the EEG signal.

Time-domain EEG signals are usually analyzed by obtaining event-related potentials (ERPs). Typically, an ERP is obtained by the multiple presentation of the same stimulus and subsequent averaging of the EEG signals aligned to the onset of the stimulus. Averaging leads to the cancellation of stimulus-unrelated noise, given that the noise has a mean of zero. ERPs obtained from auditory experiments are called *auditory evoked potentials* (e.g., Burkard et al., 2007).

The transformation of the time domain EEG signal into the frequency domain is used to analyze the spectral constitution or oscillatory content of an EEG signal (e.g. Figure 1-1B), which may or may not be related to the onset of a stimulus. In complex form, both the amplitude and phase at a certain frequency can be analyzed (e.g., relative to the onset of a stimulus). Hybrids of time and frequency representations exist in the form of time-frequency representations (TFRs), which represent the fluctuation of power and/or phase within a frequency band over time.

1.3 Neural signatures of auditory selective attention in event-related potentials

Auditory evoked potentials (AEPs) are the average neural response to an auditory stimulus obtained in the time domain of the electrophysiological signal (e.g., EEG or MEG). AEPs are obtained by the multiple repetition (i.e., trials) of the identical auditory stimulus (or stimulus category) and the subsequent averaging across the single-trial EEG signal aligned to the stimulus onset. An AEP is a sequence of deflections alternating between positive and negative halfwaves, also called components (Burkard et al., 2007). Every component can be interpreted as a representation of a neural operation along the auditory pathway (see Picton, 2013). Thus, the latency of the components allows inference on the underlying neural sources of the components (e.g. Kraus and Nicol, 2008). Early (i.e., fast) components within the first 10 ms are associated with the cochlear, the auditory nerve and the brainstem. During latencies between 10 and 80 ms (i.e., middle), the auditory cortex shows a response to an auditory stimulus. Components during latencies beyond 80 ms (i.e., slow) are mainly associated with the auditory cortex followed by frontal as well as parietal brain regions (Picton et al., 1999). Thus, in auditory cortex, the processing of the stimulus is branching, such that components in the AEP might overlap in time. Neuroanatomical models suggest that the auditory cortex itself is hierarchically organized and that information flows from core to belt areas and from caudal towards rostral regions via superior temporal regions (Hackett et al., 2014; see also Venezia et al., 2019).

Researchers use AEPs to investigate when and where attention is involved in the processing of an auditory input (for a review: Picton, 2013). Knowing the 'when' and 'where' in turn allows inference on the neural implementation of attentional filters. It is of interest whether attention acts early at a sensory stage or late on a higher order stage such as the semantic level (see above). Note that the attention-related dichotomy of early and late selection does not necessarily reflect in early and late components of an AEP, since higher order features could possibly be extracted at a relatively early stage along the auditory pathway, same as lower order features such as location could possibly affect later ERP-components as well.

It is still under debate which features of an auditory input are relevant for the neural system and which stage of the auditory pathway shows earliest signatures of auditory attention. To answer this question, in the 1960's, ERPs have been recorded (including AEPs) from human

participants undergoing selective tasks (for a review: Näätänen, 1975). Early studies on AEPs showed enhanced N1- and P2-components in the neural response to attended compared to ignored auditory stimuli (Hillyard et al., 1973), which was replicated in numerous studies (e.g., Tiitinen et al. 1993; Woldorff et al. 1993; Rimmele et al. 2011). The underlying neural generators of the N1-component were localized in auditory cortical as well as frontal and parietal brain regions (Näätänen and Picton, 1987; Picton et al., 1999).

Besides the investigation of the neural *when* and *where* of attention, researchers asked *how* attention might be neurally implemented. The characterization of attention as a filter led to the theory of two basic principles: the enhancement of an attended signal (i.e., amplification) and the inhibition of an ignored signal (i.e., attenuation, suppression). The alternation between states of high and low excitability of neurons plays an important role in this regard. Single cell recordings showed that attention enhances the firing rate (Motter, 1993) and sharpens the tuning curves of neurons sensitive to features of an attended stimulus (Spitzer et al., 1988).

The alternation between states of high and low excitability of neuronal ensembles in the primary auditory cortex was shown to align to the rhythmic structure of attended versus ignored stimuli, resulting in a neural oscillation (Lakatos et al., 2008; Bosman et al., 2009; Luo et al., 2010). Similarly, the phase of such an oscillation was shown to depend on the attended (and ignored) stimulus frequency, which showed that similar neuronal mechanisms working both on the temporal and the spectral feature dimension build up a spectro-temporal attentional filter (Fritz et al., 2007; Lakatos et al., 2013).

Spectro-temporal filtering alone would not be robust enough to broad-band distractors (see Willmore at al., 2014). Considering that broad-band noise fills temporal sound intensity troughs of an attended (speech) stimulus, the dynamic range (i.e., variance) of the attended stimulus decreases. This problem might be overcome by the adaptation of the neuronal gain in the auditory cortex, which preserves and amplifies the left variance (e.g., Rabinowitz et al., 2011).

Within this thesis, primarily late (>50 ms) evoked cortical components of the neural response to continuous speech will be discussed, since earlier research has mainly presented attention-related modulation at this stage (see below). However, recent studies have found attentional *top-down* modulation at the level of the brainstem as well (Forte et al., 2017, Etard et al., 2018).

1.4 Phase-locked neural response and the neural tracking of attended versus ignored speech

Neural phase-locking refers to the observation that firing patterns of single neurons or populations of neurons reproduce the temporal structure of a stimulus. Temporal information in acoustic signals such as speech are manifold and unfold on different time scales (Rosen, 1992; Ding and Simon, 2014).

As a framework, Rosen (1992) defined three physical features of speech signals that carry temporal information: envelope, periodicity and fine-structure. The temporal envelope is the instantaneous magnitude of a (broad-band) sound, hence, it describes the fluctuation in sound intensity. The fine structure describes the actual sound wave form containing the instantaneous phase and frequency information. Periodicity describes whether the carrier signal is rather tonal (i.e., periodic wave form) or noise-like.

Based on this framework, the relative contribution of the temporal components to speech comprehension were quantified. Besides the temporal fine-structure (Shannon et al., 1995; Obleser and Weisz, 2012), the broad-band temporal envelope moved into the focus of studies. It was shown in animals that slow and fast temporal modulations are differently represented in the auditory cortex, which led to inference on neural phase-locking to lower modulation frequencies and the temporal resolution of auditory cortical processing (Wang et al., 2003). Later, the low-frequency (1–8 Hz) cortical tracking of the speech envelope was shown in humans using magnetoencephalography (MEG; Luo and Poeppel, 2007; Ding and Simon, 2012), electrocorticography (ECoG; Mesgarani and Chang, 2012, Zion Golumbic et al., 2013) and electroencephalography (EEG, Horton et al., 2014, O'Sullivan et al., 2014).

Neural phase-locking to speech has also been called *neural tracking* of — or entrainment to — speech (Ding and Simon, 2014). Calling the whole phase-locked response to speech *entrainment* can be somewhat misleading, since it implies that there is some neural oscillator going along with the speech signal based on the temporal regularities such as the syllable rate. Such entrainment to frequency modulations was found for non-speech signals (Henry and Obleser, 2012). However, there is no clear evidence for instantaneous (i.e., non-time-lagged) or even preceding neural activation to continuous speech (e.g. Alexandrou et al., 2018). Observed neural entrainment to the envelope of speech can also be explained by a model of superimposed evoked responses,

unfolding after onsets in the envelope (Howard and Poeppel, 2010). In contrast, the term *neural tracking* implies that the neural response is following (i.e., time-lagging) the speech signal, the way it is usually observed. Hence, the term *neural tracking* will be used throughout this thesis.

Neural tracking of speech has been investigated from different methodological perspectives. A general estimate of the degree of phase-locking to a speech signal can be obtained by the calculation of inter-trial phase coherence (ITPC; Lachaux et al., 1999), which reflects the strength of alignment of neural (oscillatory) phase across multiple stimulus repetitions (trials). It was shown that the phase of slow cortical oscillations (4-8 Hz) in auditory cortical regions discriminates different spoken sentences and predicts speech intelligibility (Luo and Poeppel, 2007). Some limitations come with ITPC such as the necessity for multiple trials of the same stimulus (or at least the same temporal structure) and the restriction to the phase only, neglecting the amplitude of the neural response.

Researchers have come up with methods to describe the temporal relationship between the speech envelope and the EEG. In early approaches, the speech envelope and the EEG were simply cross-correlated within a certain range of time lags (Hertrich et al., 2012; Hambrook and Tata, 2014; Petersen et al., 2016). The goal of this approach is the detection of time-lagged similarities. Anecdotally reported, this approach however has been criticized for neglecting the auto-correlation of both speech and EEG signals.

To overcome this issue, the neural response was extracted by a regression-based method (Lalor et al., 2009; Lalor and Foxe, 2010; Crosse et al., 2016; see section 2.5 for methodological details). This method results in a *temporal response function* (TRF), which is a linear finite-impulse response model of the neural response to a certain representation of a continuous stimulus. Similar to *auditory evoked potentials*, the components of a TRF consist of multiple succeeding positive and negative deflections from zero, usually unfolding between time lags of 0 and 500 ms. A conceptual similar approach was used to obtain the stimulus-frequency-dependent temporal response to continuous speech, which was called *spectro-temporal response function* (STRF; Ding and Simon, 2012). Note that the abbreviation STRF is ambiguous with the somewhat related term spectro-temporal receptive field, which refers to the neuronal preference for certain spectro-temporal modulation patterns (Aertsen et. al, 1980; Richard et al., 1995). Unless stated otherwise, here STRF will refer to the earlier term *spectro-temporal response functions*.

Another approach to estimate the phase-locked relationship between a speech signal and the neural response is *mutual information* (MI). MI is an information-theoretic measure that captures non-parametric, non-linear relationships between two variables (Shannon, 1948; Ince et al., 2010). Based on MI, relationships between various time scales of speech signals and the neural response were observed (e.g., Keitel et al., 2018). In sum, the investigation of speech tracking has shown that predominantly temporal brain areas associated with the auditory cortex track speech signals.

The discovery of *neural tracking* of speech directly prompted researchers to ask how *neural tracking* of speech is affected by auditory attention. It was shown that a clean representation of the attended talker is predominantly represented at the stage of the auditory cortex (Mesgarani and Chang, 2012; Ding and Simon, 2012; Zion Golumbic et al., 2013; Horton et al., 2014, O'Sullivan et al., 2014). The long-standing question; At which stage of the auditory pathway attention sculptures the representation of the continuous speech out of the mixture of multiple sound sources, was now narrowed down. However, strictly controlled stimuli such as the matching of long-term sound intensity between the attended and the ignored talker did not allow one to generalize those findings to the manifold of possible listening conditions. Assuming the auditory system makes use of multiple cues in order to selectively process the input, we can expect that neural strategies to overcome challenging listening are as manifold as the listening conditions themself. Thus, the *neural tracking* of speech has to be further investigated from multiple perspectives.

1.5 Induced alpha oscillations and auditory selective attention

The observed neural responses to auditory stimuli are not only characterized by slow potentials in the delta and theta range which phase-lock to the stimulus onset (i.e., evoked oscillations), but the amplitude of a neural oscillations can be modulated as well by auditory stimulation without a concomitant phase reset. Such oscillations are referred to as induced oscillations, since stimulation is not triggering the oscillation *per se*, but rather changes the gain (e.g., Tallon-Baudry et al., 1999; Pfurtscheller, 2003; David et al., 2006). Due to its random phase, induced oscillations usually cancel out after time-domain-averaging across trials (i.e., ERPs; Wöstmann et al., 2017a). Hence, to obtain total power, time-frequency representations are calculated for single trials and subsequently averaged across trials. Time-frequency representations can be obtained by various

methods such as short-term Fourier transformation, wavelet analysis or bandpass-filtering with subsequent envelope extraction. To obtain induced power only, the ERP must be subtracted from the single trials beforehand.

Induced oscillations have been interpreted as signatures of *top-down* rather than *bottom-up* selective attention, not least because of the task-dependent dynamics of *alpha power*. For example, in an early study by Adrian (1944), *alpha power* oscillations were shown to increase when subjects switched from visual to auditory attention. Supported by subsequent studies that found enhanced *alpha power* in task-irrelevant sensory areas, alpha oscillations were interpreted as the idling rhythm of currently irrelevant sensory brain areas (Pfurtscheller and Klimesch 1992; Salmelin and Hari 1994; for review, see Pfurtscheller, 1996). Similarly, the gating-by-inhibition hypothesis postulates that *alpha power* configures brain networks by inhibition of irrelevant areas in order to guide cognitive processes (Jensen and Mazaheri, 2010). Throughout listening to speech, the modulation of *alpha power* has been related to the general demand for cognitive resources (see Weisz et al., 2011, Strauß et al., 2014). Consequently, it can be assumed that *alpha power* plays an important role for the distribution of cognitive resources in the process of selective attention.

Not only between, but also within modalities, attention-dependent *alpha power* modulation can be observed. Lateralized spatial attention in the visual (Worden et al., 2000), auditory (Kerlin et al 2010) and tactile modality (Haegens et al., 2011) leads to a hemispheric imbalance of *alpha power*. In a dichotic speech task, that the temporal synchronization of *alpha power* lateralization to the rhythm of speech predicts behavioral performance (Wöstmann et al., 2016). In sum, these studies emphasized the involvement of *alpha power* into *top-down* attentional control and strengthened the hypothesis of *alpha power* representing a distractor-inhibiting neural strategy.

1.6 Neurally steered hearing aids

Hearing aids amplify the incoming sound mixture to compensate for sensorineural, frequency-dependent hearing loss by amplification. However, increasing the sound intensity does not overcome epiphenomena such as decreased dynamic range due to an increased hearing threshold (e.g., Shapiro, 1979). Hence, additional algorithms that increase the signal-to-noise ration are necessary.

Current noise-suppression algorithms built in hearing aids rely on heuristics of stereotypical listening scenarios (for a review, see: Levitt, 2001; Bentler, 2005; Doclo et al., 2010). While the efficacy of noise reduction algorithms might be proven in the lab environment, it is not the case in highly unpredictable and versatile real-world listening scenarios. For example, broad-band noise emitted by an air condition is seldomly attended by a listener and can be suppressed by default. However, if multiple human voices are part of the incoming sound mixture, an automatic decision criterium is hard to define. Hence, a closed-loop solution that informs the hearing aid about the listener's focus of attention is necessary to overcome the limitations of hard-wired noise suppression.

Based on the findings that electrophysiological signals such as MEG and EEG are informative of a listener's focus of auditory attention (e.g., Ding and Simon, 2012; Horton et al., 2013; O'Sullivan et al., 2014), the idea emerged to feed neural information about the listener's attentional focus and/or the demand for attentional control back to the hearing aid (Lunner and Gustafsson 2014; Bleichner et al., 2015; Mirkovic et al.; 2015, Biesmans et al., 2016). Hence, a neurally steered hearing aid must first analyze the auditory scene and the neural data simultaneously and in a second step, find relations between the two that are informative of the current attentional state of the listener.

The endeavor of neurally steered hearing aids came with two main research questions: First, which signatures of auditory attention are represented in electrophysiological signals across the manifold of listening conditions? Second, which configuration of sensors is needed to capture the neural signatures of auditory attention? While answering the first question requires neurocognitive methods of measurement, the second question is more technical in nature. Nevertheless, the two research questions are tightly linked. For instance, the topographical spread of a attention-indicating neural signature directly influences the configuration of sensors needed to capture this neural signature. Hence, existing studies cannot be clearly associated with one or the other question, but rather operate in between.

It has been shown that reduced sets of EEG electrodes placed around the ear and inside the ear canal capture neural responses to both visual and auditory stimuli (Looney et al., 2012; Debener et al 2015; Bleichner and Debener 2017). Enhanced *alpha power* during eye-closure, auditory and visual steady-state responses as well as visually evoked potentials can be recorded from electrodes

placed inside the ear canal and within the pinna (Looney et al., 2012; Mikkelsen et al., 2015). EEG electrodes in the ear canal were also shown to capture neural signatures of visual selective attention such as a P300 component (Bleichner et al., 2015). Signatures of auditory attention in form of enhanced N1- and P2-components were shown with electrodes placed around the ear (Bleichner et al., 2016). In sum, those studies showed that the results from well-established EEG paradigms can be reproduced with a reduced set of electrodes (i.e., the "keyhole hypothesis"; Mikkelsen et al., 2017), albeit the signal-to-noise ratio was found to be lower compared to conventional scalp EEG.

Whether neural signatures of auditory attention to concurrent, continuous speech can be captured with a reduced set of electrodes has been investigated with electrodes placed around the ear (Mirkovic et al., 2016) and in the ear canal (Fiedler et al., 2017; see section 4.2). The studies show that even a small number of electrodes in the periphery of the ear capture the attentional modulation of brain responses to continuous speech. However, compared to conventional, multi-channel scalp EEG, the detection accuracy was found to be lowered.

Still an open question is where to place a second electrode as a reference in order to best capture the neural signatures of attention. The ideal reference electrode should capture all the noise which is also captured by the signal electrode but at the same time capture nothing of the neural signal of interest. Since scalp potentials originating from both neural and noise sources are wide-spread across the scalp, such an ideal, clear-cut reference-to-signal electrode configuration does not exist. However, the location, orientation as well as the distance between the two electrodes are crucial (Mirkovic et al., 2016; Fiedler et al., 2017; Narayanan and Bertrand, 2018; Denk et al., 2018). In sum, the studies suggest that a distance of a few centimeters is sufficient if the orientation of two electrodes placed inside or around the ear is directed towards fronto-temporal scalp regions.

The paradigmatic transition from discrete auditory stimuli such as tones or syllables to continuous speech has pushed lab-based selective listening tasks further towards real-life listening scenarios. However, the paradigms still do not capture the manifold of real-life listening scenarios. For example, the presented concurrent talkers are usually matched in sound intensity and do not move in space. Some studies have shown that the *neural tracking* of speech is robust to degradation of the attended speech signal. For example, it was shown that the *neural tracking* of a speech signal is robust to degradations caused by stationary noise down to an SNR of –6 dB

(Ding and Simon, 2013). Decoding of the attentional focus was shown to be robust to reverberation (Fuglsang et al., 2017). Interestingly, the decoding accuracy of two spatially separated talkers was shown to increase when the speech signals were filtered with *head-related transfer functions* (HRTFs) instead of presented dichotically (Das et al., 2016). In sum, those particular findings lead to the conclusion that the detection of the focus of auditory attention is possible across the manifold of listening conditions. However, this must be proven with first prototypes of neurally steered hearing aids in in real-life settings. One caveat of most of the studies named above is the fact that stimulus reconstruction methods were applied, such that a detailed investigation of the neural responses (i.e., TRFs) is not provided. This hinders conclusions on the neural strategies that lead to such a noise-robust tracking of attended speech.

As explained above, the feasibility of EEG signals to steering of a hearing aid have so far been only investigated based on the phase-locked neural responses to speech. However, it was also shown that the amplitude modulation of induced alpha oscillations is indicative of a listener's attentional state and the spatial focus of attention (Obleser and Weisz, 2012; Wöstmann et al., 2015). Alpha might be a valuable indicator of a listener's attentional focus for the steering of a hearing aid as well, especially when it comes to dynamics of the listening difficulty due to the modulation of background noise.

1.7 Research questions

The research questions of this thesis are posed between two thematic cornerstones: First, the neural implementation of auditory selective attention as its signatures can be observed in the electroencephalogram. Second, the application of those neural signatures by way of asking whether they are captured by a reduced set of electrodes in order to neurally steer a hearing aid based on a listener's focus of auditory attention.

The first part of this thesis (see chapter 3) tries to answer the question, how continuously varying listening conditions are neurally compensated for and how *bottom-up* capture of attention is avoided by the dynamic adaptation of *top-down* neural strategies. Study 1 tests whether phase-locked neural responses to concurrent speech are modulated by varying demands for *top-down* and *bottom-up* attentional control. To this end, the signal-to-noise ratio between an attended and ignored talker was varied dynamically (see section 3.1). Study 2 tests whether

induced alpha oscillations indicate the demand for enhanced *top-down* control, such that the modulation of alpha resamples the dynamically varying SNR (see section 3.2). Study 3 tests whether the lateralization of induced alpha oscillations indicates a listener's focus of spatial attention and if this lateralization interacts with the demand for *top-down* attentional control. To this end, the location of the talkers and their SNR were varied dynamically (see section 3.3).

The second part of this thesis tries to answer the question whether neural signatures of auditory attention can be captured by a reduced set of EEG electrodes placed around the ear and inside the ear canal (see chapter 4). Study 4 tests whether the neural response to auditory stimuli rich of spectro-temporal modulation can be captured at in-ear EEG electrodes (see section 4.1). Study 5 tests whether the phase-locked neural responses to concurrent tone streams and mixed speech recorded at in-ear EEG electrodes are indicative of a listener's focus of attention (see section 4.2). Study 6 tests whether the varying demand for *top-down* control reflected in the phase-locked neural responses can be captured at in-ear EEG electrodes as well (see section 4.3).

2 General methods

The general motivation for the applied methods and the underlying assumptions will be described in this chapter. More detailed methodological information can be found within the Method sections of the particular studies.

2.1 Continuous speech stimuli

Throughout this thesis, primarily audiobooks were presented to the subjects. The presentation of continuous speech instead of trial-based presentation of single words or sentences was motivated by two factors: First, in the recent years it has been shown that neural responses to continuous speech can be extracted (see section 1.4). Second, our results should contribute to the development of neurally steered hearing aids (see section 1.6). Thus, the listening scenarios presented in the laboratory should emulate real listening scenarios as far as possible. However, real-world scenarios are rarely merely listening, but rather multisensory, interactive experiences. Hence, we presented speech stimuli based on certain assumptions which go along with certain restrictions explained in the following paragraphs.

We presented two audiobooks simultaneously, which were read by professionals and are thus more predictable than free speech. However, we chose audiobooks, since this easily accessible, professionally recorded speech. We specifically chose audiobooks which were read without much excitement in order to keep their saliency constant. One has to keep in mind that a real-world conversation is much more unpredictable, since conversations are neither scripted nor one-directional. The conversational partner can switch within seconds. However, attending to one of two simultaneously presented audiobooks should challenge the basic mechanisms of auditory selective attention.

We always presented concurrent talkers of different genders (cf., O'Sullivan et al., 2014). We instructed participants to either attend to the female or male voice, while the other voice was always of the other gender, such that the cue was unambiguous. This implies that the participants could heavily rely on the spectral cues such as pitch and timbre during stream segregation. Such a clear basis for stream segregation is not always given in real-world listening scenarios, since the voices of an attended and an ignored talker can be more similar in such situations.

In most of our experiments (with exception of study 3), talkers were presented without any spatial cues (i.e., diotic listening; e.g. Cokely and Hall, 1991), which means that we presented the identical mixture of both talkers to both ears. Previous studies have predominantly presented dichotic speech, which means that different speech signals were presented to each ear respectively (Ding and Simon, 2012; O'Sullivan et al., 2014). Both, diotic and dichotic speech are extreme concepts which rarely exist in the real world. The presentation of dichotic speech may lead to an interaction between the topographical distribution of the neural response across the scalp and its modulation by attention. Since our goal was to infer on the neural response phase-locked to the temporal modulation of the talkers' speech signals, we initially avoided spatial separation. In study 3 (see section 3.3), we presented talkers continuously moving along the frontal azimuth, which was emulated via *head-related transfer functions*. In this experiment however, we only analyzed *alpha power* modulation here.

Another aspect to be considered is how to ensure that subjects are indeed following the instructions. Here, participants were asked questions regarding the content of the to-be-attended audiobook. Such an approach may invite participants into the task, but the amount of collected behavioral data is minimal and therefore does not allow an extensive analysis. Thus, this thesis mainly focusses on the effect of task instruction (i.e., attended talker an ignored talker) on the neural response.

In sum, with the presentation of continuous speech stimuli our goal was to increase the ecological validity compared to trial-based designs. However, there are still limitations left which have to be considered during interpretation of the results.

2.2 EEG sensors

Within this thesis, two kinds of EEG sensor configurations were used: First, conventional scalp EEG configurations in a 64-channel layout based on the 10-20-system (Klem et al., 1999). Second, in-ear EEG configurations consisting of three electrodes per ear canal.

The in-ear EEG devices were crafted individually for every subject and provided by *Eriksholm Research Centre* (Oticon A/S, Snekkersten, Denmark). To this end, impressions of the ear canal and outer ear were taken by trained audiologists. Based on the impressions, ear molds were 3D-printed (Figure 2-1). Subsequently, three holes were drilled into the ear molds. Electrodes were

made from a fine silver thread with a diameter of 3 mm cut into slices of 2 mm and glued into the holes. Two electrodes pointed upwards and one electrode pointed downwards with a slight orientation towards the front. The electrodes were connected to a standard 3-pin HiPro Easyfit plug.

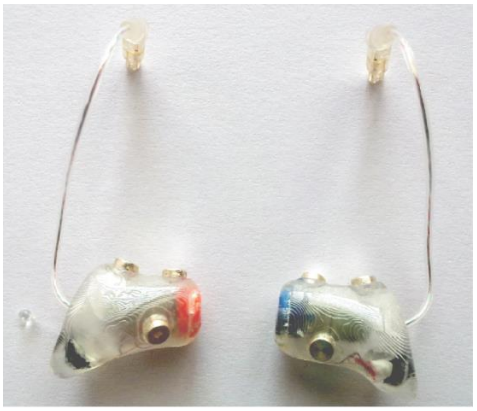

Figure 2-1: Ear molds with in-ear EEG electrodes. Ear mold were individually fitted for every subject.

2.3 Extraction of auditory features from continuous speech

Speech is transferred via fluctuations in air pressure which reach our eardrum. From there, the signal is transformed at every stage along a hierarchical structure of the auditory pathway. Henceforth, the electrophysiological representation of the speech signal is also changing in an incremental manner (e.g., Picton, 2013). Since the neural response to continuous, non-repetitive speech cannot be extracted with the ERP approach of multi-trial averaging (sees section 1.4), some representation of the speech signal must be derived in order to make inferences regarding its neural encoding. Albeit that the formulation of an overarching model of the auditory pathway up to the auditory cortex is still a work in progress (e.g., McDermott and Simoncelli, 2011; Verhulst et al., 2018), the representation of the broad-band as well as the spectrally resolved envelope in auditory cortex has been revealed by several studies (Mesgarani and Chang, 2012; Zion Golumbic et al., 2013). Nevertheless, the exact procedure of extracting the envelope of a speech signal is far away from a standardized understanding and lies to some degree in the discretion of the analyst.

The main goal of this thesis is to extract the overall neural response to continuous speech and elucidate how it is modulated by attention. Since the extraction of the neural response relies on a regression approach (see section 2.5.2), the number of regressors should be kept to minimum in order to avoid overfitting. At the same time, highly correlated regressors hinder conclusions on the most important features driving the neural response. Since the features of speech nested in the

broad-band temporal envelope are temporally correlated (see Ding and Simon, 2014), a regressor that captures the overall temporal modulation at the cost of fidelity might explain more of the neural variance than a fine-grained set of regressors such as the cochleogram.

A straightforward approximation to the broad-band temporal envelope of speech is the magnitude of the analytic signal (Figure 2-2A; e.g., O'Sullivan et al., 2014). Note that this procedure is sometimes inaccurately called Hilbert-transformation, which refers to only one of the operations within the calculation of the analytic signal (e.g., Lyons, 2004). Furthermore, auditory peripheral models of the cochlear exist and can be applied to extract a cochleogram of a sound (e.g., *NSL-toolbox*; Chi et al., 2005; *auditory modelling toolbox*; Hohmann, 2002). It was shown that the cochleogram summed across frequency leads to a better approximation of the representation of the speech envelope in the EEG (Biesmans et al., 2016). Hence, we followed the latter approach (Figure 2-2A).

Most of the above-mentioned studies investigated the neural representation of the broad-band speech envelope. However, there is strong evidence that the auditory cortex is particularly sensitive to the rate of change of the envelope (i.e., acoustic edges; onsets and offsets; Howard and Poeppel, 2010; Doelling et al., 2014, Daube et al., 2018). Especially onsets (i.e., instances of positive increase) were shown to evoke neural responses in dedicated superior temporal gyrus regions, whereas the sustained, envelope-like features were found to be tracked in middle temporal gyrus regions (Hamilton et al., 2018, see also Brodbeck et al., 2018). Importantly, it must be considered that the representations of onsets, envelope and offsets are highly cross-correlated, since every onset is followed by a peak in the envelope (here around 80 ms), which is in turn followed by an offset (here around 100 ms; Figure 2-2A & B). Interestingly, we found the strongest neural responses (TRFs) to the envelope onsets (Figure 2-2C). Most importantly, the latencies of the TRF components have the highest similarity to ERPs when the envelope onsets are used as regressors (Fiedler et al., 2016), whereas the TRFs to envelope inaccurately suggest that a neural response emerges before the presentation of a stimulus (Figure 2-2, middle). Hence, the representation of envelope onsets is used throughout this thesis. This representation was obtained by zeroing negative values of the first derivative of the envelope (i.e., half-wave rectified first derivative, Figure 2-2A).

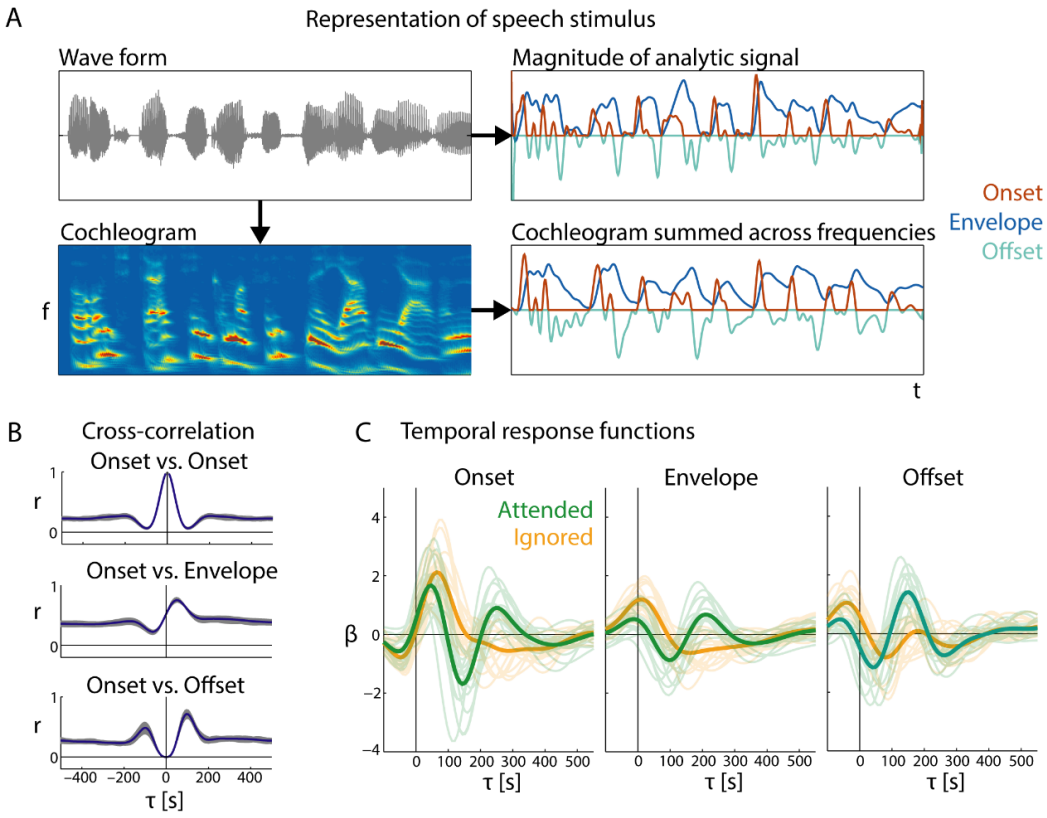

Figure 2-2: Extraction of representations of continuous speech. A) The broad-band temporal envelope can be approximated by the magnitude of the analytic signal. The transformation via a cochleogram and subsequent summing across frequencies approximates a broad-band envelope better matching the neural processing along the auditory pathway. By zeroing negative (or positive) values of the first derivative of the envelope, representations of the onsets (or offsets) can be derived. B) Onsets, envelope and offsets share similarity in a time-lagged fashion, which has been revealed by auto- and cross-correlation of the representations derived from one hour of two different audiobooks, respectively. C) TRFs to onsets, envelope and offsets share a high morphological similarity, but are shifted in time. Onsets evoke the strongest neural responses and the TRFs have the characteristics of a causal filter (i.e., weakest deflections from zero for negative time lags.).

2.4 Pre-processing of EEG data

2.4.1 Filter design

The goal of filtering electrophysiological signals is to get rid of stimulus- or task-unrelated noise such as low-frequent drifts or high-frequent noise originating from irrelevant brain areas or non-brain sources like muscles or electromagnetic interference caused by external sources. There is no one-size-fits-all filter design, but the filter must be designed according to the research question and the underlying assumptions (Widmann et al., 2014). One important consideration is whether a *finite impulse response filter* (FIR) or an *infinite Impulse response filter* (IIR) is applied and how a potential phase shifts will be compensated.

If not stated otherwise, we used high- and low-pass Hamming-window FIR-filters. For the analysis of *neural tracking* of speech, the lower and higher cutoff frequencies were set to 1 and 10 Hz, in order to avoid reduction of the amplitude within the passband between 2 and 8 Hz (the frequency range where *neural tracking* was previously observed; see section 1.4; Figure 2-3). The filter order was set to 375 (high-pass) and 100 (low-pass), respectively. On the one hand, this ensured the removal of low-frequent drifts (Figure 2-3A & D), which is important because a trial-wise baseline cannot be applied in continuous data. On the other hand, a comparably steep frequency roll-off at the higher cut-off reduced non-phase-locked alpha activity.

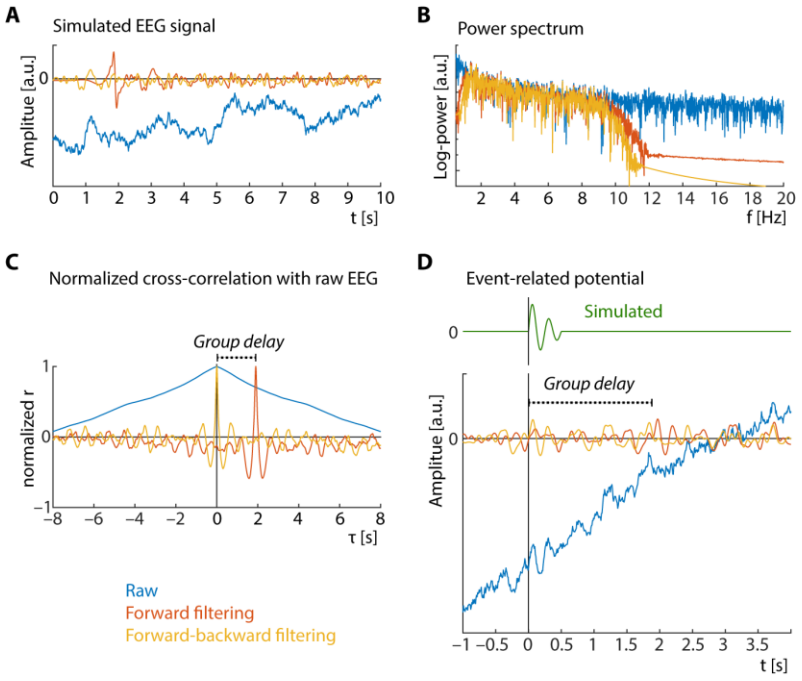

Figure 2-3: Essentials of filter design. Raw EEG signal was simulated as the cumulative sum of random gaussian noise. A) The raw EEG signal (blue) was filtered forward (red) and filtered again backward (yellow) with Hamming-window FIR filters as a low-pass (cut-off frequency f_c: 10 Hz; order: 100) and a high-pass (cut-off frequency f_c: 1 Hz; order: 375). B) Simulated raw EEG has a power spectral slope similar to real EEG signals. Filters were designed to keep the cut-off frequencies outside the pass-band between 2 and 8 Hz. C) One-directional filtering introduces a group delay of approximately 2 seconds, which can be revealed by cross-correlation between the raw and filtered EEG signal. Backward filtering compensates the group delay. D) An event-related potential was simulated as two cycles of a 4 Hz sine with decaying amplitude. Multiple instances (240 trials) of this ERP were distributed across the simulated EEG signal and averaged. Low-pass-filtering removes high-frequent noise whereas high-pass-filtering removes drifts.

The group delay of approximately two seconds was compensated by forward- and backward-filtering the EEG signal (i.e., two-pass), which doubles the effective filter order (Figure 2-3B & C). Two-pass filtering is non-causal, which means that the calculation of the current sample in time is also based on samples which lie in the future. This has to be kept in mind during interpretation of the results, because pre-stimulus deflections from zero may occur which cannot be interpreted as pre-stimulus activity *per se*, but might be due to smearing within the passband.

As exemplarily shown in Figure 2-3D, filtering removes drifts and high-frequent noise. However, noise within the passband still leads to deflections from zero not related to the ERP, which cannot be reduced by a filter but only by the acquisition of more electrophysiological data, such that stimulus unrelated noise cancels out.

2.4.2 Independent component analysis

Independent component analysis has been invented to split-up a mixture of multiple sources of variance into its underlying components (e.g., Comon, 1994). Originally invented as a method for *blind source separation* mainly applied to acoustic signals, it was later applied to electrophysiological data in order to get rid of irrelevant components such as eye blinks or muscle activity (Makeig, 2004). Importantly, the maximal number of components that can be separated equals the number of independent sensors.

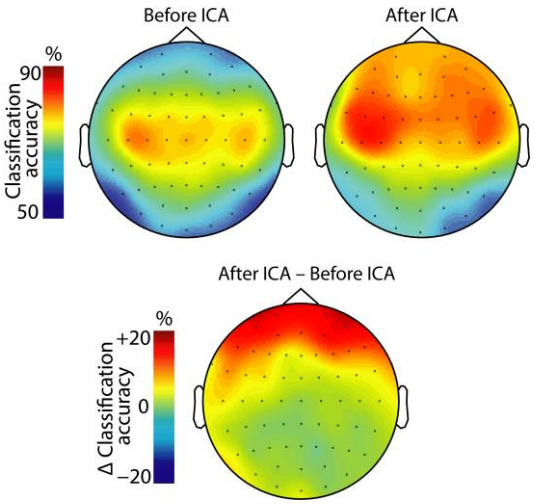

Figure 2-4: The effect of independent component analysis (ICA) on classification accuracy (i.e., *neural selectivity*). Classification accuracy is mainly enhanced at frontal channels after ICA by up to 20%, which is possibly related to the removal of noise related to eye blinks and eye movement. Electrodes close to the ear are only slightly affected by the noise removed by ICA (preliminary analysis of data of 18 subjects; see study 1) for further details.

Within this thesis, we applied ICA only on scalp EEG data (see chapter 3; study 1–3), since in those studies, we were investigating the cognitive mechanisms of selective auditory attention. In the studies about the feasibility of in-ear EEG we did not make use of ICA (cf., O'Sullivan et al., 2015), since we claimed that our results can be achieved with single EEG channels, which would not hold true if we applied ICA based on all sensors beforehand (see chapter 4; study 4–6). Interestingly, ICA mainly improves classification accuracy (i.e., *neural selectivity*; see section 2.7) at frontal EEG channels, most probably due to the removal of eye artifacts (Figure 2-4). This has implications on the single-channel classification based on electrode configuration close to the ear, which seems not as extensively affected by eye-artifacts.

2.5 Estimation of the neural response to continuous auditory stimuli

After deciding which representation of the stimulus to use, we next had to decide which method to apply in order to extract the temporal relationship between this stimulus representation and the brain signal (i.e., EEG). This requires some review of established methods in order to decide which one best fits our research questions.

2.5.1 Forward vs. backward modelling

Generally, in advance of such a modelling a crucial decision between two options must be made (Figure 2-5; see also Naselaris et al., 2011; Holdgraf et al., 2017): First, are we aiming to predict the brain signal at one sensor (or voxel) based on the representation of multiple stimulus features (e.g. talkers; Figure 2-5, right)? This is called a *forward, prediction* or *encoding* approach (also known as *forward model*). It is called *forward* because we usually assume that a brain response is following a stimulus with a certain time lag and that we can *predict* the future brain signal based on the current (and past) state of the stimulus representation.

Or second, are we aiming to reconstruct the stimulus representation based on the brain response of multiple sensors (or voxels; Figure 2-5, left)? This is called a *backward, reconstruction* or *decoding* approach (also known as *backward model*). It is called *backward* because we *reconstruct* the past stimulus representation based on the current (and future) state of the brain response.

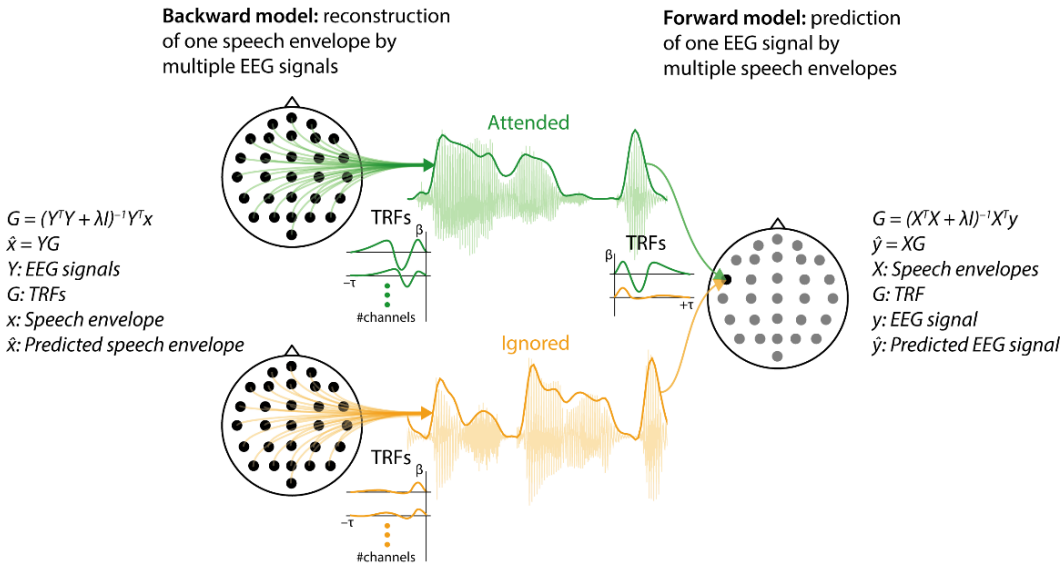

Figure 2-5: Reconstruction of one speech envelope by a *backward model* and prediction of one EEG signal by a *forward model*. Left: *Backward model* reconstruction of a single EEG envelope based on multiple EEG channels. Middle: Speech signals and envelopes of the attended and the ignored talker. Right: *Forward model* prediction of one EEG signal based on multiple speech envelopes.

Both *backward* and *forward models* have in common that the relationship between the stimulus representation and the brain signal is expressed in a kernel. Since here we are mainly investigating the temporal, time-lagged relationship between the stimulus representation and the brain signal, this kernel is called *temporal response function* (TRF; see section 2.5.2).

Mathematically, forward and *backward models* are similar, but each approach comes with strengths and weaknesses (Haufe et al., 2014). The advantage of a *forward model* is that we can directly analyze and contrast TRFs between conditions just as ERPs or even disentangle brain responses to simultaneously presented stimuli, such as attended and ignored stimuli (e.g., Jia et al., 2017). Furthermore, *forward models* allow to make inference on the topographical location of encoding in the brain. However, prediction at a single EEG channel (or voxels) does not draw on the full potential of the whole set of recorded electrophysiological data. This is where the *backward model* has an advantage, since the information of all sensors (or voxels) can be used to reconstruct the representation of one stimulus. However, since only one stimulus representation can be reconstructed at a time, a contrast between the TRFs to two stimulus categories is not directly feasible. Furthermore, an exact inference on the topographical brain location of encoding cannot be made (Haufe et al., 2014; see also Popov et al., 2018).

Within this thesis, primarily *forward models* have been applied due to two reasons: first, we wanted to investigate the attention-related effects on the morphology of the TRFs including their topographical distribution across the scalp. Second, our goal was to investigate the classification accuracy at single-channel EEG configurations and their potential to steer a hearing aid.

2.5.2 Estimation of *temporal response functions*

The *temporal response function* (TRF) reflects the temporal relationship between a stimulus representation and the electrophysiological signal, such that in can be used to predict an EEG signal in a *forward model* approach (see section 2.5.1). The basic assumption is that the EEG signal results from the convolution of the stimulus representation and the TRF plus some stimulus-unrelated noise:

$$\boldsymbol{y(t)} = \boldsymbol{x} * \boldsymbol{g} = \sum_{\tau} [\boldsymbol{x(t} - \boldsymbol{\tau)} \cdot \boldsymbol{g(\tau)}] + \boldsymbol{n(t)} \qquad \textbf{2-1}$$

where *x* is the stimulus representation, *g* is the TRF, τ is the time lag between the stimulus representation and the EEG signal y, *n* is the noise. Note that convolution is the mathematical operation to calculate the output of an FIR-filter (here the EEG) using the input signal (here the stimulus representation) and the impulse response of the filter (here the TRF). Thus, the TRF can be described as an FIR-filter, which shapes the amplitude and the phase-delay of the stimulus representation such that it becomes the EEG signal (e.g., Lalor et al., 2006). In other words, convolution means that every sample of the EEG signal is the weighted sum of past stimulus samples and the TRF is comprised of those time-lag-dependent weightings. Thus, the estimation of the TRF boils down to multiple linear regression (e.g., Fox, 2015).

Analogous to linear regression, a point of debate is how to find the TRF from some known dataset (stimulus and EEG) that best predicts an unknown EEG signal based on a stimulus. In other words: Which TRF minimizes the prediction error? To answer this question, first the measure of the prediction error must be defined (i.e., error term). The mean-squared error (MSE) is the most common target of minimization (see ordinary least squares; e.g., Fox, 2015), and is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \qquad \textbf{2-2}$$

where n is the number of observations (i.e., samples of the EEG signals), *Y* is the measured EEG signal and $\hat{Y}$ is the predicted EEG signal. Consequently, bigger deflections between the measured and the predicted EEG signal get more strongly punished than smaller deflections. However, other error terms exist such as the mean absolute error, which usually leads to increased sparsity (most of the weights are zero). However, *ordinary least squares regression* is the only approach that has an analytical solution, such that it is computationally efficient. Consequently, the TRF can be estimated by a matrix operation

$$G = (X^T X)^{-1} X y \quad \textbf{2-3}$$

where G is a matrix containing the TRFs, X is a matrix containing the stimulus representations with its sample-wise, time-lagged replications and y is the EEG signal.

A common issue in multiple linear regression is multicollinearity of the regressors, which takes effect on the estimated TRFs since neighbored samples in the TRFs are not independent due to its low-pass characteristic. This results in implausible high-frequency artifacts and edge-effects (Figure 2-6). Those artifacts can be avoided by regularization, which modulates the degree of how strongly the multicollinearity affects the weightings. Regularization can be introduced by adding the identity matrix *I* multiplied with the regularization parameter λ as follows:

$$G = (X^T X + \lambda I)^{-1} X y \quad \textbf{2-4}$$

Interestingly, regularization only slightly affects the outcome measures *neural tracking* and *neural selectivity* in a *forward model* (Figure 2-6), as it was also confirmed by a comprehensive comparison of various methods for regularization (Wong et al., 2018). Consequently, within this thesis, λ was set such that the artifacts disappeared (more details can be found in the method sections). Nevertheless, note that Wong et al. (2018) found that regularization has a stronger impact in a *backward model*.

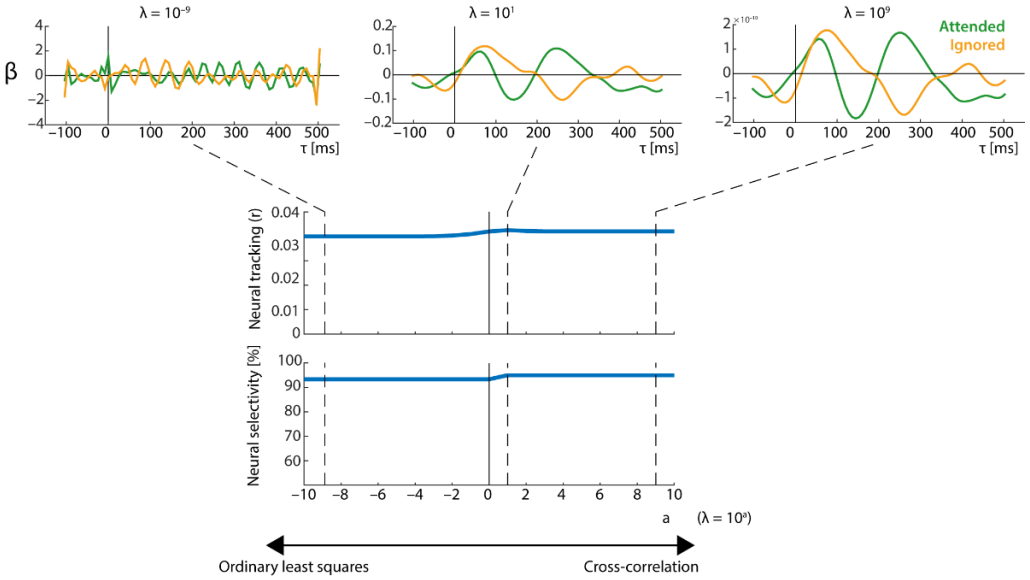

Figure 2-6: The influence of varying degrees of regularization in a *forward model* approach. *Temporal response functions* of a representative subject were estimated with different degrees of regularization. Note that the more the regularization parameter λ converges to zero, the more *ridge regression* converges to *ordinary least squares regression*. On the other side, the greater λ, the more *ridge regression* converges to cross-correlation (i.e., neglecting multi-collinearity of the regressors).

2.6 Goodness of fit as a measure of *neural tracking*

The *goodness of fit* is a measure to describe the predictive power of a model (TRFs) trained on a subset of the data and subsequently used to predict unknown EEG data based on the stimulus representation. Throughout this thesis, a *leave-one-out cross-validation* approach was chosen, which means that one part of the data, such as a block of one-minute length was left out during training of the TRF and subsequently predicted. Mathematically, the prediction of the EEG signal is the convolution of the stimulus representation with the trained TRF (Eq. 2-1). In matrix-based notation, this operation is expressed as

$$\hat{y} = XG \qquad \textbf{2-5}$$

where $\hat{y}$ is the predicted EEG signal, which is a vector since single-channel EEG signals are independently predicted (see also Figure 2-5).

Now that an EEG signal was predicted, a measure for the goodness of fit must be defined. Most commonly, the Pearson-correlation coefficient r between the predicted and the measured EEG signal is calculated. This is motivated by the fact the square of r (i.e., R^2) is proportional to the explained variance of the EEG signal. Hence, it is a measure for the least mean-squared error (MSE), which was the subject to minimization beforehand. The MSE can also be used as a measure

of the inverse *goodness of fit* (Crosse et al., 2016). However, the MSE depends on the general scale of the to-be-predicted data and hence, is not directly comparable across participants and studies.

Given that the noise *n(t)* (see Eq. 2-1) is stationary across a whole experiment, the to-be-explained variance in the EEG signal depends on the strength of *neural tracking*. In other words, the stronger the *neural tracking*, the more the recorded EEG contains signal relative to noise. Hence, the Pearson-correlation coefficient is a direct measure of the strength of *neural tracking*. Importantly, in contrast to the most commonly used case of Pearson correlation, negative coefficients do not have the same meaning as positive coefficients. For example, a value of −0.5 does not mean that the prediction is as good as a prediction that results in a value of 0.5 (even if they result in the same R^2), since we expect that the predicted and the measured EEG signal have the same polarity and that negative correlation coefficients can only be caused by noise. A Pearson correlation coefficient of around 0.05 was typically observed in previous studies (e.g., O'Sullivan et al., 2014), which means that the predicted EEG signals explains 0.25% of the variance of the recorded EEG signal, which corresponds to a signal-to-noise ratio of approximately −60 dB.

2.7 Classification accuracy as measure of *neural selectivity*

Within the recent decades, the field of machine learning has increasingly gained popularity due the increasing availability of data and the exponentially growing computational resources (Waldrop, 2016). One main objective of machine learning is classification of unknown data. To this end, data is usually categorized in classes, such that the goal of classification is to predict the membership of a data point to a class (e.g., Kotsiantis et al., 2007). Classification accuracy refers to the percentage (or proportion) of correctly classified trials. The field of machine learning has brought up a myriad of approaches, describing which, would go beyond the focus of this thesis. Loosely defined, classification is a multi-dimensional, non-parametric statistical test of the difference between the mean of two classes, where classification accuracy is the outcome measure.

Here, a classification approach is applied to detect a listener's focus of auditory attention. The two classes can be labelled as *attend talker A* or *attend talker B* and one data point belongs to a certain time frame (e.g., a block of five minutes where the female talker was attended). The classification is based on the coefficients obtained by Pearson correlation between the predicted EEG signal and the measured EEG signal as described above (see section 2.6). Crucially, not only

one EEG signal is predicted based on the trained TRFs, but a second EEG signal is predicted by flipping the labels of the TRFs (i.e., *attended* and *ignored*). Subsequently, both predicted EEG signals are Pearson-correlated with the measured EEG signal. The two resulting correlation coefficients are compared. One of the two is representing the class *attend talker A* and the second is representing *attend talker B.* Since we instructed participants to attend to one of the talkers, the Pearson correlation coefficient that corresponds to this instruction should be more positive. If this is the case, the classification is correct. The percentage of correctly classified trials will be referred to as classification accuracy, while the chance level is 50%.

Since the prerequisite for yielding a classification accuracy above chance is a numerical difference between the TRFs to the attended and the ignored talker, it directly expresses how differently and how consistently attention shapes the morphology of the TRFs. In other words, classification accuracy indicates neural selective processing. Throughout this thesis, classification accuracy will be referred to as *neural selectivity*.

2.8 Overview of experiments

For greater comprehensibility, all conducted experiments are listed below (see Table 1). Given that multiple experiments were analysed both within a single study as well as in various studies, an overview is provided to increase the comprehensibility.

Exp.	N	Stimuli	Task	EEG Sensors	Study
1	7	Concurrent Oddball Concurrent Audiobooks	Attend left/right Attend male/female	In-ear/ Scalp	5
2	6	Natural sounds	One-back task	In-ear	4
3	18	Concurrent Audiobooks with modulated SNR	Attend male/female	Scalp	1,2
4	6	Concurrent Audiobooks with modulated SNR	Attend male/female	In-ear / scalp	6
5	25	Concurrent Audiobooks with modulated SNR and location	Attend male/female	Scalp	3

Table 1: Overview of experiments

3 Neural adaptation to continuously varying acoustic conditions

3.1 Study 1: Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions[1]

3.1.1 Abstract

Listening requires selective neural processing of the incoming sound mixture, which in humans is borne out by a surprisingly clean representation of attended-only speech in auditory cortex. How this *neural selectivity* is achieved even at negative signal-to-noise ratios (SNR) remains unclear. We show that, under such conditions, a late cortical representation (i.e., *neural tracking*) of the ignored acoustic signal is key to successful separation of attended and distracting talkers (i.e., *neural selectivity*). We recorded and modelled the electroencephalographic response of 18 participants who attended to one of two simultaneously presented stories, while the SNR between the two talkers varied dynamically between +6 and −6 dB. The *neural tracking* showed an increasing early-to-late attention-biased selectivity. Importantly, acoustically dominant (i.e., louder) ignored talkers were tracked neurally by late involvement of fronto-parietal regions, which contributed to enhanced *neural selectivity*. This *neural selectivity*, by way of representing the ignored talker, poses a mechanistic neural account of attention under real-life acoustic conditions.

3.1.2 Introduction

Human listeners comprehend speech surprisingly well in the presence of distracting sound sources (Cherry, 1953). The ubiquitous question is how competing acoustic events capture *bottom-up* attention (e.g., by being dominant, that is, louder than the background), and how in turn *top-down* selective attention can overcome this dominance (e.g., listening to a certain talker against varying levels of competing talkers or noise; Kaya and Elhilali, 2017).

Auditory selective neural processing has been mainly attributed to auditory cortex regions. It is by now well-established that the auditory cortical system selectively represents the (spectro-)

[1] *This section is highly adopted from a published article (Fiedler et al., 2019) with contributions to the study design, analysis and writing from Malte Wöstmann, Sophie K. Herbst and Jonas Obleser.*

temporal envelope of attended, but not ignored speech (i.e., neural phase-locking; Magneto-encephalography: Ding and Simon, 2012; Electroencephalography: Kerlin at al., 2010; Power et al., 2012; Horton et al., 2013; O'Sullivan et al., 2014). Accordingly, auditory cortical responses allow for a reconstruction of the spectrogram of speech and to detect the attended talker (e.g., Mesgarani and Chang, 2012; Zion Golumbic et al., 2013). In sum, selective neural processing in auditory cortices establishes an isolated and distraction-invariant spectro-temporal representation of the attended talker.

However, as has been shown, degradations of the acoustic signals attenuate the neural phase-locking to speech. Experimental degradations have included artificial transformations of temporal fine structure (Ding et al., 2014; Kong et al., 2015), rhythmicity (Kayser et al., 2015), reverberation (Fuglsang et al., 2017) or decreased signal-to-noise ratio (SNR; Kong et al., 2014; Ding and Simon, 2013; Giordano et al., 2017). Not least, neural selection of speech appears weakened in people with hearing loss (Petersen et al., 2016). In sum, those studies suggest that the strength of neural phase-locking indicates behavioral performance such as speech comprehension.

Additionally, higher order non-auditory neural mechanisms facilitate speech comprehension as well. The supra-modal, fronto-parietal attention network is a candidate to be involved in *top-down* selective neural processing during demanding listening tasks (Woolgar et al., 2016). Beyond the phase-locking in lower frequency bands (i.e., approx. 1 – 8 Hz; Wang et al 2018, Pomper and Chait 2017), *top-down* selective neural processing has also been associated with changes in the power of induced alpha-oscillations (i.e., approx. 8 – 12 Hz; Obleser and Weisz 2012; Kayser et al. 2015, Wöstmann et al. 2016). Specifically, increased parietal alpha-power is related to enhanced suppression of the distracting input (Wöstmann et al., 2017b). This reflects that, besides the neural spectro-temporal enhancement of the attended talker, a crucial role in *top-down* neural selective processing was attributed to the suppression of the ignored talker.

Neural signatures of suppression can be two-fold. First, suppression can attenuate the neural response to an ignored talker compared to an attended talker, like it was found in neural phase-locking from latencies of around 100 ms (Ding and Simon, 2012; Wang et al., 2018). Second, active suppression can add or increase components in the neural response to the ignored talker, given that the response is dissociable from the response to the attended talker (e.g.; a louder ignored talker evoking a stronger neural response anti-polar to the response to a louder attended talker).

Here we asked, how the components of the phase-locked neural response are affected by selective attention under varying signal-to-noise ratio (SNR).

The phase-locked neural response to broad-band continuous speech can be obtained from EEG by estimating the (delayed) covariance of the temporal speech envelope and the EEG, which results in a linear model of the cortical response; a *temporal response function* (TRF; Lalor et al., 2009; Crosse et al., 2016). Analogous to the *event-related potential* (ERP), the components of the TRF can be interpreted as reflecting a sequence of neural processing stages where later components reflect higher order processes within the hierarchy of the auditory system (Davis and Johnsrude, 2003; Picton et al., 2013; Di Liberto et al., 2015).

Here, we use a listening scenario in which two concurrent talkers undergo continuous SNR variation. Our results demonstrate differential effects of *bottom-up* acoustics vs. *top-down* selective neural processing on earlier vs. later neural response components, respectively. Source localization reveals that not only auditory cortex regions are involved in the selective neural processing of concurrent speech, but that a fronto-parietal attention network contributes to selective neural processing through late suppression of the ignored talker.

3.1.3 Methods

3.1.3.1 Participants

Eighteen native speakers of German (9 females) were invited from the participant database of the *Department of Psychology*, *University of Lübeck*, Lübeck, Germany. We recruited participants who were between 23 and 68 years old at the time of testing (mean: 49, SD: 17), to allow valid conclusions from such a challenging listening scenario to middle-aged and older adults. All reported normal hearing and no histories of neurological disorders. Incomplete data due to recording hardware failure was obtained in four more, initially invited participants. All participants gave informed consent and received payment of 8 €/hour. The study was approved by the local ethics committee of the University of Lübeck.

3.1.3.2 Stimuli

The goal of this study was to investigate the selective neural processing of one of two talkers under a continuously varying signal-to-noise ratio (SNR). Here, the signal is a to-be-attended

talker and the noise is a to-be-ignored talker. Our study was conducted in a within subject 2 by 3 design (attention by SNR (three levels)).

We selected two audiobooks read by native German speakers, one female (Elke Heidenreich, 'Nero Corleone kehrt zurück', read by Elke Heidenreich) and one male (Yuval Noah Harari, 'Eine kurze Geschichte der Menschheit', read by Jürgen Holdorf). The following steps of stimulus preparation were done using custom code written in *MATLAB* (Version 2017a; Mathworks Inc., Natick, MA, United States). Sequences of silence longer than 500 ms were truncated to 500 ms to avoid long parts of silence (O'Sullivan et al., 2014). The first hour of each audiobook was selected for further preparation. The first 30 minutes of each audiobook served as the to-be-attended and the rest served as the to-be-ignored speech, such that all subjects could attend both stories from the beginning and attended (and ignored) both the female and the male voice the same amount of time.

The identical mixture of the attended and ignored talker was presented on both ears, resulting in a concurrent listening scenario without any spatial cue (i.e. diotic, Figure 3-1A). Hence, the only cues available for talker segregation consisted in the spectro-temporal features of the talkers, such as pitch, formants, and amplitude modulation.

The SNR was modulated symmetrically around 0 dB. An SNR of 0 dB refers to concurrent talker signals with a matched long-term root-mean-square (rms) amplitude as used previously in numerous studies (e.g. Power et al., 2012; O'Sullivan et al., 2014; Mirkovic et al., 2015). Coming from an SNR of 0dB, the SNR was either increased to +6 dB by raising the sound pressure level (SPL) of the to-be-attended talker by 6 dB or decreased to −6 dB by raising the SPL of the to-be-ignored talker by 6 dB. Thus, the talkers were either balanced (Figure 3-1B, black) or one of the talkers was dominant (purple) and the other was non-dominant (grey). The particular SNR-range (−6 to +6 dB) was chosen to create a challenging but at the same time solvable listening task. Even if an SNR of −6 dB is rare in real-life listening scenarios (Smeds et al., 2015), the *neural tracking* of attended speech has been reported as intact at SNRs as low as −6 dB (Ding and Simon, 2013). However, speech perception (number of words repeated correctly) of normal hearing subjects starts to suffer around an SNR < 0 dB and the speech-reception threshold (i.e. 50% correct) usually lies between −5 and 0 dB (Pichora-Fuller et al., 1995, Bentler et al., 2004).

As building blocks for SNR modulation, we created a sample of *plateaus* (i.e., constant SNR of –6, 0 or +6 dB) and *ramps* (i.e., transition between *plateaus*). The length of *plateaus* was uniformly distributed between 5 and 9 seconds in discrete steps of one second. The *ramps* were linear interpolations between SNRs with the length uniformly distributed between 1 and 5 seconds in discrete steps of one second. The length distributions of *plateaus* and *ramps* were kept uniform within each talker and within their assignments as being attended or ignored. We concatenated *plateaus* via *ramps* such that a 0 dB *plateau* was either followed by a +6 dB or a 6 dB *plateau*, whereas a +6 dB or a –6 dB *plateau* were always followed by a 0 dB *plateau* via a respective *ramp*. Randomly varying SNR time courses were created for each subject individually in order to avoid systematic overlap between the SNR modulation and the audiobooks. Stimulus material was cut into twelve blocks, which resulted in an average block length of five minutes. Sound files were created with a sampling rate of 44.1 kHz and a 16-bit resolution. The experiment was implemented in the software *Presentation* (*Neurobehavioural Systems*, Berkeley, United States). Stimuli were presented via headphones (*HD25*, *Sennheiser,* Wedemark, Germany).

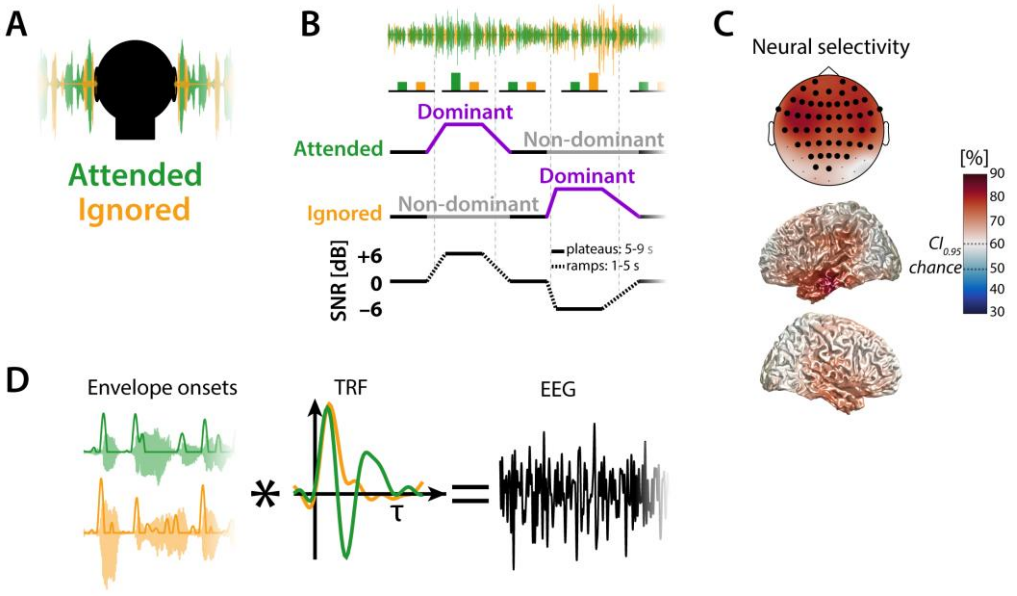

Figure 3-1: Experimental design, *forward model*, and *neural selectivity*. A) Two mixed talkers (female & male) were presented on both ears without spatial segregation (diotic). B) The signal-to-noise ratio (SNR) between attended (signal) and ignored (noise) talker was varied between –6, 0 and +6 dB by either raising the level of the attended talker or the ignored talker. Length of *ramps* and *plateaus* were drawn from uniform distributions. C) *Neural selectivity*, here expressed as classification accuracy in detection of the attended and ignored talker, were averaged across subjects. Shown here is accuracy as obtained by prediction of EEG signals (Fiedler et al., 2017) at single EEG channels and single voxels in source space, respectively. Highlighted channels of topographic maps indicate that the lower bound of the confidence interval (bootstrapped mean on the group level) was greater than the 95%-confidence bound of a binomial distribution ($CI_{0.95}$ = 60%; Combrisson and Jerbi, 2015). D) *Temporal response functions* (TRF) to the attended and ignored talker were extracted by a *forward model* based on the assumption that the measured EEG signal is the superposition (convolution) of the envelope onsets (of the attended and ignored talkers) and the TRFs, respectively. TRFs reflect the neural response evoked by a single envelope onset.

3.1.3.3 Task

The twelve blocks were presented such that subjects were instructed to attend to the female or to the male talker in an alternating fashion. After instruction before each block (i.e. attend to female or attend to male), subjects were asked to start the stimulus presentation by a button press, which enabled the participants to take a break between blocks. During listening, subjects were asked to fixate a cross presented on the screen in order to reduce eye movement.

Every other block, the stories picked up at the point it ended two blocks before. After each block, subjects were asked to rate the difficulty of maintaining attention by mouse-clicking on a continuous color bar ranging from red (difficult) to green (easy). For later analysis, the continuous color bar was discretized into ten segments (1 = difficult, 10 = easy). Subsequently, participants were asked to answer four multiple-choice questions concerning the content of the to-be-attended audiobook. The average rating of difficulty was neither significantly correlated with the number of questions correctly answered (Pearson's $r = 0.1$, $p = 0.73$), nor with participants' age (Pearson's $r = -0.17$, $p = 0.51$). Furthermore, we found no significant correlation of the number of correctly answered questions with age (Pearson's $r = -0.11$, $p = 0.65$).

3.1.3.4 Data acquisition and preprocessing

EEG was recorded with 64 electrodes *Acticap* (*Easycap*, Herrsching, Germany) connected to an *ActiChamp* amplifier (*Brain Products*, Gilching, Germany). EEG signals were recorded with the software *BrainVision Recorder* (*Brain Products*) at a sampling rate of 1 kHz. Impedances were kept below 10 kΩ. Electrode TP9 (left mastoid) served as reference during recording.

The EEG data were pre-processed in *MATLAB 2017a* (*The MathWorks, Inc.*, Natick, Massachusetts, United States) using both the Fieldtrip-toolbox (version: 20170321; Oostenveld et al., 2011) and custom-written code. The EEG data were re-referenced to the average of the electrodes TP9 and TP10 (left and right mastoids) and resampled to fs = 125 Hz. The continuous EEG data were highpass-filtered at fc = 1 Hz and lowpass-filtered at fc = 30 Hz (two-pass Hamming window FIR, filter order: 3fs/fc).

From the continuous EEG data, we extracted the parts during which the twelve blocks of audiobooks were presented (see above). For every subject, we applied independent component analysis (ICA; Makeig et al., 2004) on the concatenated data of the twelve blocks and manually

rejected components that were clearly related to eye movements, eye blinks, muscle artifacts, heartbeat as well as singe-channel noise. On average, 26 of 62 components (SD: 7.3) were rejected.

For further analysis, we lowpass-filtered the data again at fc = 10 Hz (two-pass Hamming window FIR, filter order: 3fs/fc), which assured that the amplitudes at all frequencies up to 8 Hz were not reduced. Previously, neural activity phase-locked to the envelope was only found up to a frequency of approximately 8 Hz (Zion Golumbic et al., 2013; Ding et al., 2014). We could confirm this finding by incrementally raising the cutoff frequency, which didn't change the morphology of the TRFs (see below) but only decreased the prediction accuracy due to the interference of non-phase-locked neural activity and external noise in higher frequencies.

3.1.3.5 Extraction of envelope onsets

A temporal representation of the acoustic onsets, further called envelope onsets, was extracted from the presented speech signals (Fiedler et al., 2017). Those representations later served as regressors to model neural responses to the talkers (see below). First, we extracted an auditory spectrogram containing 128 spectrally resolved sub-band envelopes of the speech signals logarithmically spaced between approximately 90 and 4000 Hz using the NSL toolbox (Chi et al., 2005). Second, the auditory spectrogram was summed up across frequencies, which resulted in broad-band temporal envelopes of the audiobooks. Taking the derivative of the envelope and zeroing all values smaller than zero (Hertrich et al., 2012) returned the envelope onsets, which only contain positive values at time periods of an increasing envelope, as can be found at acoustic onsets (Figure 3-1C).

Using the envelope onsets as regressor does not imply that we only modeled the encoding of acoustic onsets. Every onset is followed by a peak in the speech envelope (Figure 3-1C), which is then again followed by an offset and the next onset and so forth, resulting in a high autocorrelation between those features. Nevertheless, onsets are the earliest feature that could possibly evoke a neural response (Picton, 2013). The latency of modeled responses to envelope onsets (compared to envelopes) was found to be most similar to conventional ERPs (Fiedler et al., 2017, supplemental material).

3.1.3.6 Estimation of temporal response functions

We applied an established method to estimate a linear forward (encoding) model (Lalor et al., 2009; Crosse et al., 2016). The model contains *temporal response functions* (TRFs), which are estimations of the neural response to a continuously varying stimulus feature. In our case, this stimulus feature is the envelope onsets (see above) of both, the attended and the ignored talker. Based on the assumption that every sample in the EEG signal r(t) is the superposition of neural responses to past onsets and thus can be expressed for one talker by a convolution operation:

$$\boldsymbol{r(t) = s * TRF = \sum_{\tau}[s(t-\tau) \cdot TRF(\tau)]} \qquad \textbf{3-1}$$

where s(t) is the envelope onsets, TRF is the *temporal response function* that describes the relationship between s and r over a range of time lags τ (Figure 3-1C). The TRF contains a weight for each time lag τ. We investigated time lags in the range from –100 to 500 ms. In order to obtain the β-weights of the TRF to both talkers contained in the matrix G_{TRF}, ridge regression (Hoerl and Kennard, 1970) was applied, which can be expressed in the linear algebraic form:

$$\boldsymbol{G_{TRF} = (S^T S + \lambda m I)^{-1} S^T R} \qquad \textbf{3-2}$$

where *S* is matrix containing the envelope onsets of both the attended and ignored talker and its sample-wise time-lagged replications, *R* contains the measured EEG signal, λ is the ridge parameter for regularization, the scalar *m* is the mean of the trace of S^TS (Biesmans et al., 2016) and *I* is the identity matrix.

Note that the usage of the scalar m is equivalent to normalizing the auto-correlation matrix X^TX (i.e., dividing it by the variance of the regressors), such that amount of regularization is proportional to the variance of the regressors.

The optimal ridge parameter λ was estimated according to Fiedler et al. (2017) and was set to λ = 10. Due to the low-pass characteristic of the envelope onsets, we need regularization because neighboring samples are highly co-linear. Using the time-lagged envelope onsets as regressors, this co-linearity usually results in high-frequent artifacts as well as implausible high regression weights at the edges of the TRFs (see section 2.5.2). We iteratively increased λ and inspected the TRFs. We chose the lowest λ that removed those high-frequent artifacts.

TRFs were estimated on a trial-by-trial basis, where trial refers to a part (e.g. a *plateau* of +6 dB) of certain length cut from the continuous stimulus and the respective EEG data. For the subsequent analysis, we subdivided the data in two ways: First, to get a general estimate of the model's ability to dissociate between attended and ignored talkers, we cut the data into one-minute trials, resulting in trial lengths comparable to previous studies (O'Sullivan et al., 2014; Mirkovic et al., 2015; Biesmans et al., 2016; Fiedler et al., 2017). This resulted in 60 trials per subject. Second, we cut the data based on the applied SNR modulation, which resulted in three groups of trials: −6 dB, 0 dB and +6 dB. To use the entire recording, the data were cut at the time points where *ramps* of the SNR time courses either crossed −3 dB or +3 dB (Figure 3-1B). This resulted in 180 trials of 0 dB and 90 trials of −6 and +6 dB, respectively. The average length of those trials was 10 seconds (i.e. average length of a *plateau* (7 seconds) and average length of two halves of a *ramp* (2x1.5 seconds)). In order to balance the number of trials across SNRs, 90 trials from 0 dB were randomly drawn from the 180 trials of every subject. During the analysis, we contrasted TRFs not only within conditions, but also contrasted the TRFs to the talkers within their role of being *dominant* (Fig 2B, purple; attended under SNR = +6 dB, ignored under SNR = −6 dB) or *non-dominant* (Fig 2B, grey; attended under SNR = −6 dB, ignored under SNR = +6 dB). We will use those terms and schematic bar graphs (Figure 3-1B) throughout the entire section.

3.1.3.7 Statistical analysis on temporal response functions

To extract significant spatio-temporal deflections in the TRFs at an SNR of 0 dB, we applied a two-level statistical analysis (two-level cluster-test; e.g. Obleser et al., 2012). At the single-subject level, we used one-sample t-tests to test the TRF to the attended, the ignored as well as the attended–ignored difference against zero. Resulting t-values were transformed to z-scores. At the group level, the deflection of z-scores from zero was tested by a cluster-based permutation one-sample t-test (Maris and Oostenveld, 2007), which clusters t-values with p-values < 0.001 of adjacent time-electrode bins (with a minimum of 4 neighboring electrodes). The extracted cluster is compared to 4,000 clusters drawn randomly from the data by permuting condition labels. The resulting cluster p-value reflects the relative number of Monte Carlo iterations in which the summed t-statistic of the observed cluster is exceeded. This contrast indicates how components of the TRF are generally affected by attention under balanced conditions.

In a second step, the identical cluster-based permutation test was applied to obtain significant differences between the TRFs depending on whether a talker was *dominant* or *non-dominant*. This contrast was separately computed for the attended an ignored talker and it indicates, how the TRFs are affected by changing SNR.

In a third step, the difference between the TRFs to the attended and ignored talker were contrasted separately for *dominant* and *non-dominant* talkers. This contrast describes how attention affects the TRF to a *dominant* talker (easy-to-attend, hard to ignore) or a *non-dominant* talker (hard-to-attend, easy-to-ignore), respectively.

For illustration of the neural responses, we averaged single-subject TRF β-weights across channels of interest. Channels of interest were defined as the channels being part of both significant clusters found in the attended–ignored difference between TRFs under a balanced SNR of 0 dB (Figure 3-2B). The 95%-confidence-bands were obtained by bootstrapping (Efron, 1979) across the averaged TRFs of all subjects, using 4,000 iterations.

3.1.3.8 Neural tracking and neural selectivity

To further disentangle *bottom-up* and *top-down* effects, we investigated the TRFs based on two measures: *neural tracking* and *neural selectivity*. While *neural tracking* is a measure of how strongly a talker is encoded in the EEG (irrespective of attention), *neural selectivity* is a measure of how differential (i.e., attended vs. ignored) those representations are due to the impact of selective attention.

As a base for those two measures, we followed the forward method of predicting EEG signals and comparing those to the measured EEG signal, as described in detail by Fiedler et al. (2017). In a leave-one-out fashion, we predicted EEG signals of a single trial contained $\hat{R}$ in following the equation:

$$\widehat{\boldsymbol{R}} = \boldsymbol{S}\boldsymbol{G}_{\boldsymbol{TRF}} \qquad \textbf{3-3}$$

where S is the matrix containing the envelope onsets and G_{TRF} is the matrix containing the trained TRFs.

Neural tracking was defined as the Pearson-correlation coefficient r between the predicted and recorded EEG signals using the estimated TRFs (see above). While TRFs are zero-centered and

components alternate between positive and negative deflections like ERPs, one advantage of the r-value is its directionality towards positive values (negative values are due to noise). Hence, the strength of *neural tracking* can be directly evaluated without dissociation between positive and negative deflections.

Neural selectivity was defined as the percentage of trials the TRFs could successfully identify a talker as being attended or ignored. Therefore, two different EEG signals were predicted per trial (Eq. 3), the first representing a talker being attended and the second representing the same talker being ignored. While one of the EEG signals is representing the task instruction (i.e., attend the to-be-attended talker; ignore the to-be-ignored talker), the other EEG signal represents the alternative (i.e. attending the to-be-ignored talker; ignoring the to-be-attended talker). We calculated the Pearson correlations for both predicted EEG signals with the measured EEG signal (Fiedler et al., 2017). Talker identification was successful if the EEG signal referring to the task instruction yielded higher correlation. Note that during unbalanced SNRs (i.e., −6 dB & +6 dB), the alternative EEG signal was predicted based on the TRFs estimated on the opposite SNR (e.g., under an SNR of +6 dB, the alternative to attending the to-be-attended talker (*dominant*) is ignoring the to-be-ignored talker under an SNR of −6 dB).

Since this is a *forward model* approach, *neural tracking* and *neural selectivity* were obtained at every single EEG channel (Crosse et al., 2016). Likewise, both measures were obtained at the source level at every single voxel. We split up the prediction by either using only the prediction of the to-be-attended, only the prediction of the to-be-ignored or the sum of both predictions, such that the talker-specific contribution to *neural tracking* (*neural selectivity*) could be compared to the overall *neural tracking* (*neural selectivity*).

In order to evaluate the unfolding of *neural tracking* and *neural selectivity* over TRF time lags, we used a sliding-window of time lags (size: 48 ms, 6 samples) with an overlap of 24 ms (3 samples) for the prediction. For every position of the window, *neural tracking* and *neural selectivity* were calculated (see above).

To estimate if the found effects are just random observations, we created surrogate data by circularly shifting the stimulus relatively to the EEG signal, such that the temporal structure was preserved but the stimulus-to-EEG relationship in time got distorted. The number of shifted

samples was randomly varied and at least one second (125 samples). This procedure was done during prediction of every single trial, such that we obtained the same amount of values for surrogate *neural tracking* and *neural selectivity*. As a result, we obtained 95%-confidence bands for both the observed and the surrogate *neural tracking* and *neural selectivity*.

In advance of any arithmetic operation on *neural tracking*, the underlying Pearson-correlation coefficients were fisher-z-transformed. Accordingly, *neural selectivity* (i.e., percentage correct) was logit-transformed.

3.1.3.9 Source localization

To further trace the origin of effects observed in sensor space, we applied LCMV-beamforming (Drongelen et al., 1994; Van Veen et al., 1997) to obtain source-activity time courses in single voxels of the brain. Using a standard template brain from Fieldtrip/SPM (Montreal Neurological Institute) together with the *Acticap* electrode layout, leadfields were calculated with a grid resolution of 10 mm. Individual LCMV-filter weights were obtained using 5% regularization. The continuous time-domain EEG data were projected to source space, resulting in three source activity time courses (X-Y-Z) per voxel. In order to obtain a single time course for each voxel, the direction of highest variance was determined by *principal component analysis* and used for further analysis. All further processing steps in source space were done analogously to sensor space EEG data. Note that the source maps must be interpreted with caution due to the limited spatial resolution of EEG data. We only provide source maps to support our interpretations of the significant effects found in sensor space (e.g., Sohoglu et al., 2012).

3.1.4 Results

We asked participants to listen to one of two simultaneously presented audiobooks under varying signal-to-noise ratio (Figure 3-1A & B; –6 to +6 dB SNR). After each of twelve five-minute blocks, subjects were asked to rate the difficulty of listening to the to-be-attended talker on a color bar ranging from red (difficult = 1) to green (easy = 10). The average difficulty ratings strongly varied between subjects (mean: 5.2, SD: 2.2, range: 2.3–8.9). No difference in difficulty ratings for listening to the female versus the male talker was found (one-sample t-test, $t_{17} = 1.17$, $p = 0.26$).

To test their successful attending, participants were asked to answer four multiple-choice questions on the content of the to-be-attended audiobook after each five-minute block. The

percentage of correctly answered questions was far above chance (25%) for all participants (mean: 81%, SEM: 2%, range: 60–96%). All participants were thus able to follow the to-be-attended talker.

3.1.4.1 Neural selectivity

To obtain a general estimate of which EEG channels and which voxels reveal signatures of *neural selectivity*, we identified the attended (and the ignored) talker by forward prediction of EEG signals based on one-minute parts of the EEG and envelope onsets (see methods). Overall *neural selectivity* was highest (up to 80%) at fronto-central electrodes and respective temporal cortex regions in source-space (Figure 3-1C).

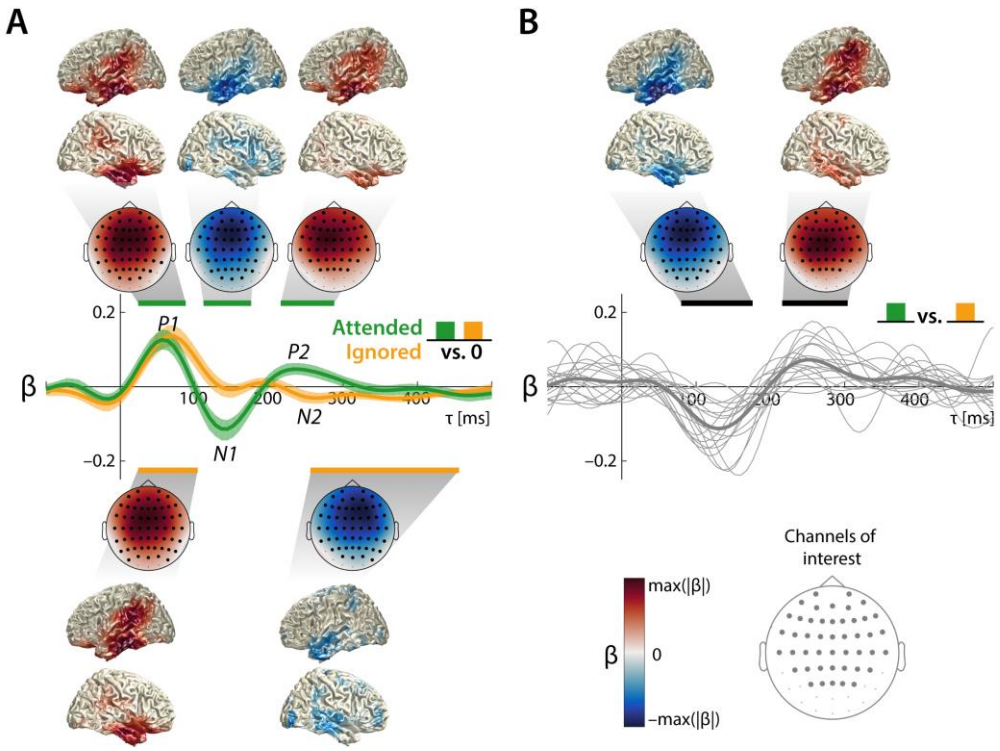

Figure 3-2: *Temporal response functions* (TRF) to continuous speech of concurrent talkers under balanced SNR (0 dB). TRF β-weights depict average across subjects and average across channels of interest. Confidence bands (95%) were obtained by bootstrapping the mean across subjects. Horizontal lines indicate time ranges of significant difference from zero obtained from a cluster-based permutation test at the group level. Topographic maps show β-weights of clusters averaged across the cluster time range. Highlighted channels are part of the significant clusters. Source localizations show the 20% most strongly contributing voxels with full opacity and faded to transparency towards zero. A) Response to the attended talker (green, upper topographic maps) clearly show a cascade of three components ($P1_{TRF}$–$N1_{TRF}$–$P2_{TRF}$). Response to the ignored talker (red, lower topographic maps) only show a $P1_{TRF}$, whereas the $N1_{TRF}$ and $P2_{TRF}$ are suppressed. B) Significant differences between neural responses to the attended and ignored talker are present in the $N1_{TRF}$- and $P2_{TRF}$-timerange. Thin grey lines show single subject TRFs averaged across channels of interest.

3.1.4.2 Attention modulates neural responses to concurrent speech

Next, we assessed in greater detail the unfolding of attentional selection of to-be-attended speech in time. To this end, we estimated the TRFs from the balanced SNR trials of 0 dB (i.e. independent of the SNR manipulation) and assessed the most prominent response components and their modulation by attention. We inspected both the TRFs to the attended and ignored talker individually (Figure 3-2A), as well as the difference between the TRFs to the attended and ignored talker (Figure 3-2B) to examine signatures of *neural selectivity*.

First, an early positive component (termed $P1_{TRF)}$ appeared in the TRFs to the attended (Figure 3-2A, 24–88 ms, $p = 2\times10^{-4}$) and ignored (Figure 3-2A, 24–112 ms, $p = 2\times10^{-4}$) talkers, but without any attention-related difference (Figure 3-2B). Latency, polarity, and topography of this component compared well to a P1 as found in auditory evoked potentials (AEPs).

Second, a later negative deflection (termed N1TRF) was only present in the TRF to the attended talker (Figure 3-2A; 112–176 ms, $p = 5\times10^{-4}$). This component was significantly increased in magnitude (i.e., more negative) for the attended versus the ignored talker (Figure 3-2B, 80–176 ms, $p = 5\times10^{-4}$). Noteworthy, the significant attentional modulation of this component (attended–ignored) started already at a time lag of 80 ms, when both the TRF to the attended and to the ignored talkers were still in positive deflection (see Figure 3-2A).

Third, a positive deflection between 200 and 300 ms (termed $P2_{TRF}$; Figure 3-2A, 216–304 ms, $p = 5\times10^{-4}$), was again only present in the TRF to the attended talker. This component mainly drove the significant difference between the responses to the attended and ignored talker (Figure 3-2B, $p = 2\times10^{-4}$).

Interestingly, in the same time interval, a negative deflection was found in the TRF to the ignored talker (termed $N2_{TRF}$; Figure 3-2A, 248–424 ms, $p = 2\times10^{-4}$). While at earlier stages, TRFs to the attended and the ignored talker showed the same polarity ($P1_{TRF}$), at the stage of the $P2_{TRF}$ we see an anti-polar relationship. Effectively, this also enhanced the late, attended–ignored difference in the $P2_{TRF}$ time range (Figure 3-2B).

In sum, three prominent components ($P1_{TRF}$, $N1_{TRF}$, $P2_{TRF}$; Figure 3-2A) were identifiable with notable consistency across individual subjects. The latter two components were absent in the TRF to the ignored talker and thus indicated *neural selectivity*. All three components ($P1_{TRF}$, $N1_{TRF}$,

$P2_{TRF}$) mainly localized to superior and inferior temporal regions (Figure 3-2A). Note that the source localizations of the two latter components ($N1_{TRF}$, $P2_{TRF}$) compared well to the sources of enhanced *neural selectivity* between attended and un-attended talkers (Figure 3-1C).

3.1.4.3 Late representation of ignored talker enhances towards more detrimental SNRs

Next, we analyzed the impact of a varying SNR on the t*emporal response functions* (TRFs). To this end, we first contrasted the TRFs of the two extreme conditions (SNRs −6 vs. +6 dB; Figure 3-3A & B). Second, we contrasted TRFs across SNRs matched for the acoustic properties of being either the louder or the quieter talker (Figure 3-3C & D), such that the occurring differences between the TRFs to the attended and the ignored talker can solely be related to *top-down* attending versus ignoring. For simplicity, we will use the terms *dominant* (attended talker under +6 dB SNR, ignored talker under −6 dB SNR) and *non-dominant* (attended talker under −6 dB SNR, ignored talker under +6 dB SNR). We observed an SNR-dependent latency shift which hindered time-lag-wise attended−ignored contrasts within SNRs.

Importantly, two later additional components appeared whenever the ignored talker was *dominant* (Figure 3-3B): the first (160−178 ms, $p = 0.04$) localized to temporal regions, while the second extended markedly into parietal regions (232−280 ms, $p = 0.001$). The enhanced involvement of parietal regions differentiated this detrimental-SNR, ignored-speech component from all others. Visual inspection of the TRFs to *dominant* talkers (Figure 3-3C) highlights the additional late N2 component in the TRF to the ignored talker, which appears to be anti-polar to the $P2_{TRF}$ to the attended talker.

In contrast, TRFs to *non-dominant* talkers (Figure 3-3D) suggest that the observed attention-related differences are decreased (cf., Figure 3-3C) due to smaller deflections of the $N1_{TRF}$ and $P2_{TRF}$ to the *non-dominant* attended talker and the lack of the anti-polar $N2_{TRF}$ to the *non-dominant* ignored talker. We summed the magnitude of the attended−ignored difference across all time lags, which revealed a smaller attended−ignored difference for *non-dominant* versus *dominant* talkers ($t_{17} = 3.80$, $p = 0.0014$). Thus, the neural response to a *dominant* ignored talker does not resemble the neural response to a *dominant* attended talker by capturing *bottom-up* attention. Instead, *dominant* ignored speech retains a distinct "ignored" neural signature, most likely to due to *top-down* neural signaling of its to-be-ignored status.

In sum, our findings indicate that, when a talker is *dominant*, neural signatures of selective processing are enhanced (compared to *non-dominant*). Importantly, this enhancement is not only affecting the representation of the attended talker, but an important contribution to this enhanced *top-down* processing can be attributed to an additional late component ($N2_{TRF}$) in the neural response to the ignored talker. To further disentangle the contribution of the selective processing of the attended and ignored talker, we established the time lag and talker resolved measures *neural tracking* and *neural selectivity*, which will be discussed in the following section.

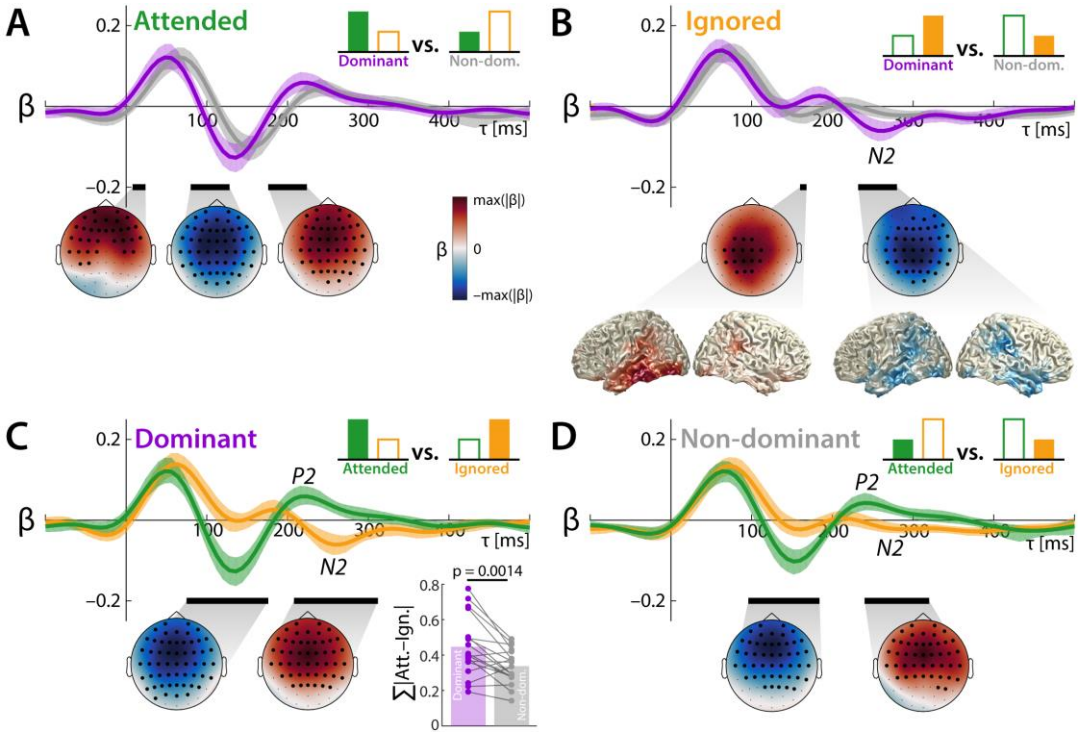

Figure 3-3: *Temporal response functions* (TRF) to continuous speech of concurrent talkers contrasted as *dominant* vs. *non-dominant* talkers and attended vs. ignored talkers, respectively. TRF β-weights depict average across (N = 18) subjects and average across channels of interest. Confidence bands (95%) were obtained by bootstrapping the mean across subjects. Schematic bar graphs indicate the investigated contrast. Black horizontal lines indicate time ranges of significant difference obtained from a cluster-based permutation test at the group level. Topographic maps show β-weight differences of clusters averaged across the cluster time range. Highlighted channels are part of the significant clusters. Source localizations show the 20% most strongly contributing voxels with full opacity and faded to transparency towards zero. A) Responses to the *non-dominant* attended talker are delayed compared to the *dominant* attended talker. B) A late component appeared in the response to the *dominant* ignored talker, which involved parietal regions. C) Late negative response ($N2_{TRF}$) to the *dominant* ignored talker appears anti-polar to the response to the *dominant* attended talker. Inset: Magnitude of the attended–ignored TRF difference summed across all time lags for *dominant* and *non-dominant* talkers. D) *Non-dominant* talkers show significant but decreased attention-related differences.

3.1.4.4 Neural selectivity increases by way of a late cortical representation of ignored speech

We established two measures to quantify the encoding and the selective neural processing of the talkers during the unfolding of the neural response reflected in the TRFs. First, *neural tracking* reflects the strength of representation (i.e., encoding) of a talker in the EEG and is related to TRF deflections from zero. Second, *neural selectivity* quantifies how accurately an attended talker can be identified as attended and an ignored talker as ignored, respectively. Thus, *neural selectivity* reflects the TRF difference between the attended and the ignored talker.

Parallel inspection of *neural tracking* and *neural selectivity* allowed us to consolidate our findings and to further disentangle the effects of *bottom-up* and *top-down* attention on the TRFs. A prerequisite for *neural selectivity* is *neural tracking* (i.e., TRF deflection from zero) of at least one of the talkers. In turn, *neural tracking* does not necessarily mean that *neural selectivity* is involved, since both talkers can be identically tracked (i.e., TRFs show same deflection from zero). Consequently, mere enhancement of *neural tracking* due to acoustic changes indicates pure increase of *bottom-up* auditory encoding, as the neural processing of both talkers is equally affected. In contrast, the co-occurrence of enhanced *neural tracking* and enhanced *neural selectivity* indicates increased *top-down* attentional selection, as the neural processing differs between the talkers.

For example, the increased sound pressure level of a to-be-ignored talker may increase its saliency and thus *bottom-up*-pull attention towards it. This would result in enhanced *neural tracking* of the ignored talker and the neural response would become less distinct from the respective response to a *dominant*, but intentionally attended talker. However, if there exists a counter-acting, *top-down* process that enhances and maintains a neural-response differentiation between the attended and the ignored talker, *neural selectivity* would increase at the same time.

To get a total estimate of *neural tracking* of the two talkers, we first used all time lags of the TRFs (i.e., –100-500 ms). Figure 3-4A shows the *neural tracking* of the attended, the ignored as well as the overall *neural tracking* of the two talkers (attended & ignored). The overall *neural tracking* was found to be well above zero for all participants as well as the *neural tracking* of the two talkers separately (Figure 3-4A, bottom).

In a next step, we estimated the time-lag- and channel-dependent unfolding of *neural tracking*. Importantly, we found enhanced *neural tracking* of the attended talker compared to the ignored talker under the balanced SNR of 0 dB (144–288 ms, $p = 0.02\times10^{-2}$, data not shown), driven by fronto-central channels. This is congruent with the time ranges and topographies of the $N1_{TRF}$ and $P2_{TRF}$, which were absent in the TRF to the ignored talker. Accordingly, enhanced *neural tracking* of the attended compared to the ignored talker was also found in the unbalanced conditions (Figure 3-4B; –6 dB: 120–192 ms, $p = 0.004$; +6 dB: 144–288 ms, $p = 0.001$).

Interestingly, towards more adverse SNRs (*dominant* ignored talker), the late enhanced *neural tracking* of the attended talker compared to the ignored talker seems to shrink (Figure 3-4B). Visual inspection of the time-lag resolved *neural tracking* suggests that this shrinkage is due to an additional late cortical representation of the ignored talker that appears when the ignored talker is *dominant*. The contrast of the *neural tracking* of the *dominant* and the *non-dominant* ignored talker confirmed such a late cortical representation (Figure 3-4C, 240–312 ms, $p = 1.5\times10^{-3}$) originating mainly from fronto-parietal as well as temporal regions.

Importantly, the overall *neural selectivity* is not affected by adverse conditions (Figure 3-4E, grey bars, –6 vs +6 dB, one-sample t-test, $t_{17} = 0.24$, $p = 0.81$). However, the relative contribution of the *neural selectivity* of the attended talker and ignored talker changes across SNRs (–6 vs +6 dB; one-sample t-test; attended: $t_{17} = -4.6$, $p = 2.77\times10^{-4}$; ignored: $t_{17} = 2.18$, $p = 0.044$): Towards more adverse SNRs, the *neural selectivity* of the ignored talker increases, while the *neural selectivity* of the attended talker decreases (Figure 3-4E, top). This is also discernible in single subjects (Figure 3-4E, bottom), where *neural selectivity* of the attended talker is stronger under an SNR of +6 dB (right, 16 of 18 subjects) and stronger for the ignored talker under an SNR of –6 dB (left, 11 of 18 subjects).

If the increased *neural tracking* of the *dominant* ignored talker at later stages (Figure 3-4C) is solely driven by its increased saliency (i.e., higher dominance evoking a stronger response), we would expect no concomitant increase in *neural selectivity* (see above). However, we found a late increase in *neural selectivity* for the *dominant* compared to the *non-dominant* ignored talker (Figure 3-4G, 216–264 ms, 2.5×10^{-3}). Neural sources compared well to the increased fronto-parietal *neural tracking* of the *dominant* ignored talker (see Figure 3-4C & G).

Furthermore, *neural tracking* and *neural selectivity* (for *dominant* vs *non-dominant* ignored speech) were positively correlated (Figure 3-4D, r = 0.78, p = 0.014×10^{-2}): If a listener's *neural tracking* was relatively strong for the *dominant* versus *non-dominant* ignored talker, the neural response allowed more accurate identification of the ignored talker as ignored.

In sum, at later stages, not only increased selective neural processing of the attended talker but also the selective neural processing of the ignored talker facilitates input segregation under adverse listening conditions.

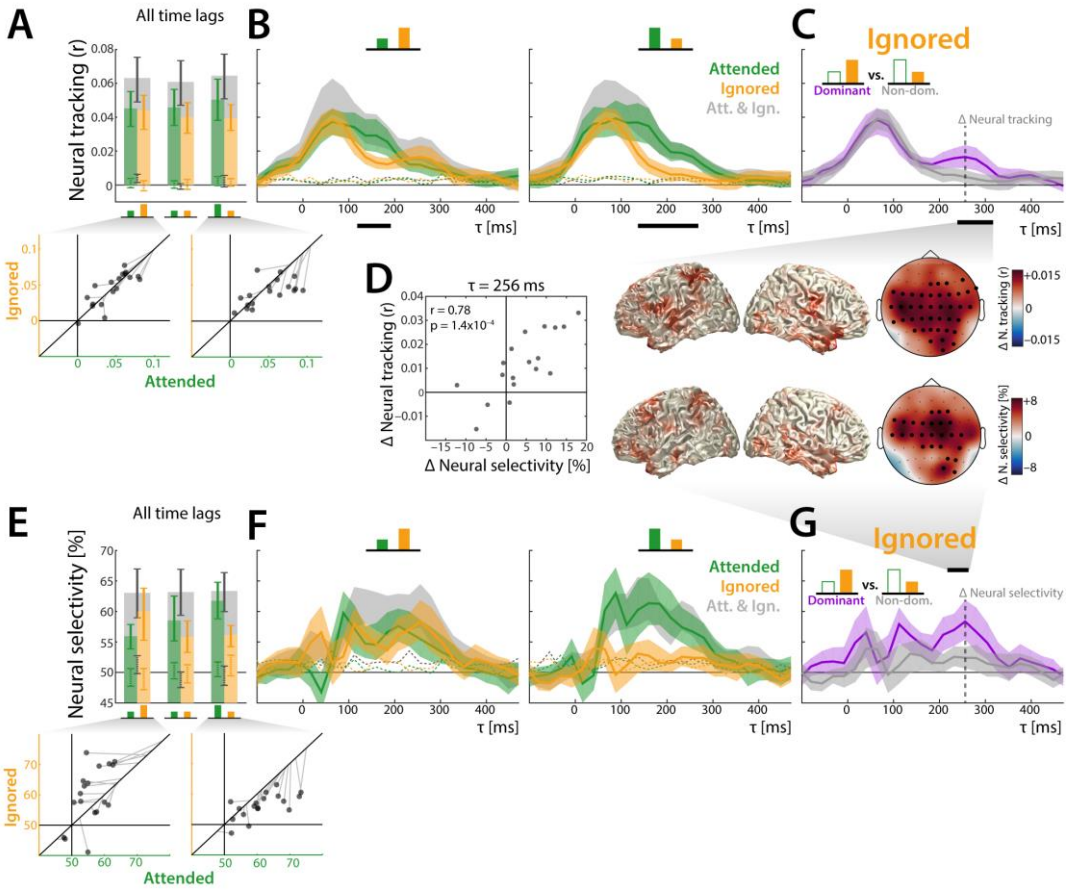

Figure 3-4: Unfolding of *neural tracking* and neural selectivity reveals late neural selective processing of the ignored talker. *Neural tracking* and *neural selectivity* were estimated based on the extracted TRFs to the attended (green), the ignored (orange), as well as both talkers (grey). Confidence bands (95%) were obtained by bootstrapping for both the observed data (solid lines and error bars) and surrogate data (dotted lines and error bars). Over time lags, only the upper confidence bound is shown for the surrogate data. Highlighted channels (topographic maps) are part of a significant cluster. Source localizations show the 20% most strongly contributing voxels with full opacity and faded to transparency towards zero. A) *Neural tracking* across all time lags (–100-500 ms). Scatterplots (bottom) show single-subject data averaged across channels of interest. Grey lines indicate overall *neural tracking* of both talkers at the 45°-line. B) Unfolding of *neural tracking* across time lags under SNR of –6 (left) and +6 dB (right). Black horizontal lines indicate time ranges of significant clusters for the difference between attended and ignored talkers. C) Contrast of *neural tracking* between the *dominant* and *non-dominant* ignored talker. D) Correlation of change in *neural tracking* and change of *neural selectivity* at τ = 256 ms. E) *Neural selectivity* across all time lags (–100-500 ms). Scatterplots (bottom) show single-subject data averaged across channels of interest. Grey lines indicate overall *neural tracking* of both talkers at the 45°-line. F) Unfolding of *neural selectivity* across time lags under SNR of –6 (left) and +6 dB (right). G) Contrast of *neural selectivity* between the *dominant* and *non-dominant* ignored talker.

3.1.5 Discussion

In the present study, human listeners attended to one of two concurrent talkers under continuously varying signal-to-noise ratio (SNR). We asked to what extent a late cortical representation (i.e., *neural tracking*) of the ignored acoustic signal is key to the successful separation of to-be-attended and distracting talkers (i.e., *neural selectivity*) under such demanding listening conditions.

Forward modeling of the EEG response revealed neural responses to the temporal envelopes of individual talkers and their modulation by both, *top-down* attentional set, and *bottom-up* SNR. Critically, towards more adverse SNRs, an additional late negative component occurred in the neural response to the ignored talker. Under adverse conditions, this component was found to be accompanied by enhanced selective neural processing (*neural selectivity*), emerging primarily from fronto-parietal brain regions.

The present result suggests that irrelevant, to-be-ignored acoustic inputs are not simply absent from the late cortical response but become actively suppressed in regions beyond auditory cortex.

3.1.5.1 Early and late neural signatures of selective neural processing

Generally, we replicated previous results that showed that attention–ignored differences in the neural response can primarily be found at time lags > 80 ms, which were mainly attributed to stronger *neural tracking* caused by enhanced N1 and P2 components in the response to the attended vs. ignored talker (Horton et al. 2013; O'Sullivan et al., 2014; Ding and Simon, 2012). Here we show that a P2-counter-acting response to the ignored talker enhances the attended–ignored difference as well.

While earlier studies showed that selective neural processing in auditory cortices is mainly working out a clean representation of the attended talker (Mesgarani and Chang, 2012; Zion Golumbic, 2013), we show that a late neural representation of a distracting auditory input is accompanied with enhanced selective neural processing in a cocktail-party scenario as well. This additional late neural representation was revealed by going beyond strictly matched sound pressure levels of attended versus ignored speech (cf., Horton et al., 2013; O'Sullivan et al., 2014, Ding and Simon, 2012; Mirkovic et al., 2015; Biesmanns et al., 2016), by presenting speech signals both as target and distractor (cf., Ding and Simon, 2013) and by applying SNR-variation

symmetrically around 0 dB (cf., Kong et al., 2014). In sum, our design allowed us to draw conclusions on the neural selective processing of real-world listening scenarios of dynamically varying listening demand.

Our investigation of concurrent speech under varying SNR helps to disentangle neural mechanisms of *early* and *late selection* (Treisman 1964). Since the ignored talker predominantly masks the attended talker under adverse listening conditions (i.e., negative SNRs, which we have labelled *dominant*), early neural filters tuned to the spectro-temporal properties of the attended talker might not be sufficient (i.e., neural gain, Willmore et al., 2014).

Thus, a later filter on the ignored signal must actively suppress distracting inputs. We found such a neural filter mechanism (Figure 3-4C & G) active in a time range which was previously attributed to processing of phonological (Di Liberto et al. 2016, Brodbeck at al. 2018) as well as semantic features (Broderick at al. 2018), which both go beyond basic acoustic properties of speech (Obleser and Eisner, 2009). One suggestion of our results is that when phonemes (or even words) of the *dominant* ignored talker pull *bottom-up* attention, their representation is actively suppressed at a late stage in order not to impair linguistic representation of the attended talker's speech.

3.1.5.2 Late distractor suppression in a non-auditory, fronto-parietal attention network

Previously, it has been shown that neural selective processing of concurring auditory stimuli is mainly accomplished in auditory cortex, resulting in a 'clean' and distraction-invariant representation of the attended talker (Mesgarani and Chang 2012; Zion Golumbic 2013).

Critically, under the adverse SNR of –6 dB, our analysis revealed an enhanced response to the ignored talker in a later time range (i.e., 200–300 ms) consisting of a positive and a negative component (Figure 3-3B). The latter is anti-polar to the $P2_{TRF}$ (to the attended talker). This additional component, which we interpret as a signature of active suppression of the ignored talker, involved non-auditory regions, which are part of the fronto-parietal attention or global-demand network (Woolgar et al., 2016), where we found enhanced neural selective processing of the ignored talker.

Under the assumption that such active suppression is costly to the cognitive system, it has been suggested that it is only deployed if necessary (Chait et al., 2010). Neural signatures for active suppression of irrelevant signals during late (~200 ms) AEPs have been examined before (Melara et al., 2002; Chait et al., 2010). Pomper and Chait (2017) related enhanced centro-parietal activity in the theta band (4–7 Hz) to enhanced *top-down* control. Parietal activity in the theta-band was also found to be inversely related to the delta-band auditory entrainment in superior temporal gyrus (Keitel et al., 2017). Here we show how late *top-down*, fronto-parietal neural processing of the distracting auditory input is unfolding in time and might facilitate overall selective neural processing.

In earlier studies, researchers highlighted the predominant tracking of the attended talker (Mesgarani and Chang, 2012; Ding and Simon, 2012, Zion Golumbic, 2013, O'Sullivan et al. 2014), emphasizing that a clean representation of the attended talker is key to successful listening. In some contrast to this, previous results shed light on the neural processing of the ignored talker (see also Wöstmann et al., 2017b, Olguin et al., 2018). We have shown here that the overall neural selective processing is surprisingly robust against such demanding listening conditions (Ding and Simon, 2013), and that a 'clean' or isolated tracking of the ignored talker is at least as essential.

This finding invites some speculation on the neural implementation of attentional filters more generally. On the one hand, a selective neural filter can be solely optimized to let pass relevant features of attended signals. On the other hand, it can be optimized to let pass features of the ignored talker, which might be relevant for suppression at a later stage. In line with earlier studies, we found that the *neural tracking* was dominated by the attended talker (speaking for the first strategy). However, under most demanding listening conditions (i.e., negative SNR), *neural selectivity* was dominated by the ignored talker.

Neural filter mechanisms might thus adapt depending on the listening demand. Follow-up studies should investigate the relationship of such filter adaptation to the concept of listening effort (Rönnberg et al., 2013; McGarrigle et al., 2014): Additional tracking of the ignored talker leads to higher neuro-computational load and might also be related to working memory performance (Rudner et al. 2011).

Within our design, we can only draw limited conclusions on the behavioral impact of the late *neural tracking* of the ignored talker. This is due to the tradeoff between sufficient behavioral data (e.g., trial-based design) and ecological validity (e.g., presentation of continuous speech; Hamilton and Huth, 2018). Following studies should acquire more fine-grained behavioral data, ideally without losing much of the ecological validity.

Our results show that, within the hierarchy of the central auditory pathways, the cocktail-party problem might look solved or settled at the stage of secondary auditory cortex (Mesgarani and Chang, 2012), but higher-order, attentional networks and their dedicated processing of distracting speech appear key to this solution.

3.1.5.3 Conclusions

The present data show how components of the unfolding *temporal response function* as identified in a *forward model* of the electroencephalographic signal can reflect distinct neural stages of attentional filtering. These stages contain the initial, attention-independent encoding of acoustic signals; the extraction and amplification of relevant features; and lastly a robust, purely attention-driven selective response to the attended and ignored acoustic signals.

Most consequential to our thinking about attentional filtering in the central auditory system, an active-suppression response to ignored acoustic signals originates from non-auditory, fronto-parietal attentional networks. In sum, with a design closer to real-life listening scenarios, our study provides insight into how selective neural processing of attended speech unfolds and is upheld not only by auditory cortices. Instead, establishing a clean cortical representation of the attended talker as suggested previously hinges on achieving a late suppression of ignored signals, with contributions by regions of the fronto-parietal attention network.

3.2 Study 2: Neural selective processing is more strongly reflected in phase-locked responses than the modulation of *alpha power*

3.2.1 Abstract

Alpha power is a dominant neural oscillation in the human brain. Across sensory modalities, *alpha power* has been associated with inhibition of task-irrelevant brain regions or neural pathways. It is also modulated by the degree of acoustic degradation of both target and distracting

auditory stimuli. Consequently, *alpha power* has been associated with the *top-down* attentional distribution of cognitive resources and a link to the concept of *listening effort* has been established. Here, we investigated if *alpha power* modulation was key to selective neural processing of concurrent and continuous auditory inputs. We recorded and modelled the time-frequency electroencephalographic response of 18 participants who attended one of two simultaneously presented talkers, while the signal-to-noise ratio (SNR) between the two talkers varied stochastically. We hypothesized that *alpha power* tracks the SNR and thus reflects the demand for inhibition of distracting auditory inputs. Our results show that there is no statistically significant modulation of *alpha power* driven by the SNR. Exploratorily, we modeled the response to various derivates of the SNR, which did not reveal statistically significant SNR-related modulation. A direct comparison between alpha band *neural selectivity* and phase-locked *neural selectivity* highlighted the prominence of *neural selectivity* in the phase-locked responses. Our results show that conclusions drawn from studies (using trial-based designs) did not directly transfer to the presentation of continuous speech. This has implications on the understanding of neural mechanisms involved in auditory selective processing as well as the development of neurally-steered hearing aids.

3.2.2 Introduction

Alpha waves are prominent oscillatory components of around 10 Hz. The *alpha power* peak usually sticks out of the otherwise 1/f-shaped power spectrum of human EEG (see section 1.2). Its saliency led researchers to the investigation of its functional role. While earliest observations showed that eye closure leads to enhancement *alpha power* (Berger, 1929), it was also observed that *alpha power* decreases when subjects were involved in a visual compared to an auditory task (Adrian, 1944). Those findings already suggested that *alpha power* is related to the distribution of cognitive resources deployed across different modalities, whose sensory areas constantly compete for attention.

Primarily in studies of the visual modality, *alpha power* has been associated with inhibition of task-irrelevant sensory input. The inhibition-timing hypothesis (Klimesch et al., 2007) is based on the observation that a decrease in *alpha power* follows the onset of task-relevant stimuli (event-related desynchronization; ERD; Pfurtscheller and Aranibar, 1977) and an increase of *alpha power* follows the onset of task-irrelevant stimuli (event-related synchronization, ERS, Worden et al.,

2000). Similarly, Jensen and Mazaheri (2010) conceptualized the "gating by inhibition" hypothesis, which proposes that the neural signal is guided through the brain by inhibition of irrelevant neural pathways. Likewise, the same mechanism might underly the suppression of an irrelevant input to avoid distraction of task-relevant areas. Hence, this mechanism is closely linked to selective attention.

Interestingly, observed *alpha power* dynamics also show a spatial component, resulting in interhemispheric *alpha power* imbalance (i.e., lateralization) depending on the task-relevant or -irrelevant location (Worden at al., 2000; Sauseng et al., 2005). It is still debated whether this lateralization is driven by both an increase of *alpha power* in the hemisphere where the distractor is processed (i.e., ipsilateral to target) and a decrease in the hemisphere where the target is processed (i.e., contralateral to target). In sum there is evidence that the dynamics of *alpha power* orchestrate hemispheric states of excitation and inhibition.

In the auditory modality, similar findings underpinned the inhibitory role of *alpha power*. Parietal *alpha power* lateralization was observed during auditory spatial attention tasks (Kerlin et al., 2010; Frey et al. 2014). Successful spatial attention was linked to enhanced synchronization of *alpha power* lateralization to the presentation rate of speech (Wöstmann et al., 2016). This suggests that *top-down* inhibitory *alpha power* supports auditory selective processing at a comparably early, temporo-spatial processing stage.

Beyond the modulation by temporo-spatial attention, *alpha power* was found to be differently modulated by the spectral degradation of attended vs. ignored speech (Obleser and Weisz, 2012; Wöstmann et al., 2017b), which suggests that *alpha power* is not only inhibitory, but more generally related to the degree of deployed cognitive effort. One aspect of deployed cognitive effort is working memory load (Rabbitt, 1968; Wingfield et al., 2005; Piquado et al., 2010), of which *alpha power* may be an indicator for. In auditory tasks, various studies show that higher working memory load is associated with increased *alpha power* (Leiberg et al., 2006; Karrasch et al., 2004). Crucially, it has been shown that the degree of spectro-temporal degradation of speech and the amount of working memory load drive similar neural processes associated with oscillations in the alpha range (Obleser et al., 2012).

Consequently, *alpha power* has been suggested to indicate listening effort (Rönnberg at al. 2013; McGarrigle et al., 2014; Peelle, 2018). The concept of listening effort still lacks a precise definition, but it refers to the amount of deployed cognitive resources to understand speech or, more generally, to solve a listening task. It was shown that *alpha power* is modulated by the degree of background noise and the amount of memory load, which indicates that more adverse listening conditions are compensated by the deployment of more cognitive resources (Petersen et al., 2015). Interestingly, Petersen et al. (2015) found that the modulation of *alpha power* is limited, which was interpreted as a signature of limited cognitive resources. Besides the phase-locked neural response to speech (which reflects the neural selection of concurrent speech mainly in sensory areas, see section 1.4), *alpha power* may indicate how much cognitive effort is deployed in order to achieve such *neural selectivity*. The mostly fronto-parietal occurrence of listening-effort-related *alpha power* modulation suggests that a more attention-general, multimodal hierarchy such as the fronto-parietal attention network may be involved.

Here, we have investigated whether *alpha power* modulation is key to selective neural processing of concurrent and continuous speech. We continuously varied the adversity of the listening condition by either raising the sound pressure level of the attended (signal) or the ignored (noise) talker, such that effectively, the signal-to-noise ratio varied over time. Since a more negative SNR leads to increased masking of the to-attended talker, we hypothesized that the varying SNR also varies the degree of deployed cognitive resources, such that a co-modulation of *alpha power* can be observed.

3.2.3 Methods

This is a re-analysis of published data (Fiedler et al. 2019). The experimental design, data acquisition and preprocessing were described in detail above (see section 3.1.3.1–3.1.3.3). The estimation of the neural response in time-frequency domain is described below in detail.

In brief, 18 subjects listened to one of two simultaneously presented audiobooks. The level of either the attended or the ignored talker was increased ubpredictably by 6 dB, which resulted in random fluctuations of the SNR between −6 and +6 dB. The 64-channel EEG data were band-pass filtered between 1 and 30 Hz and artifacts were removed by independent component analysis (ICA; Makeig et al., 2004). The EEG data were cut into blocks of approximately five-minute length according to the presentation, which resulted in twelve blocks per subject.

3.2.3.1 EEG time-frequency representation

The time-frequency representation (TFR) of the pre-processed EEG data was estimated by Morlet wavelets (e.g., Bruns, 2004; Cohen, 2014). The frequencies of interest were logarithmically spaced between 1 and 32 Hz and the number of cycles was set to seven. This resulted in a total of 31 bands (6 bands per octave) with an approximate overlap in band-width of 60% (Figure 3-5A). Along the time axis, wavelets were shifted in steps of 20 ms, resulting in a temporal resolution of 50 Hz.

Other than EEG data in the time domain, TFR values are neither zero-centered nor normally distributed, but power values are positive definite and highly skewed (Figure 3-5B). In trial-based designs, this skew is compensated by relative or dB-change baselines. Since a baseline is not applicable to continuous data, the TFR was transformed based on the equation

$$\boldsymbol{TFR_{trans}} = (\boldsymbol{TFR^{p}} - \mathbf{1})/\boldsymbol{p} \qquad \textbf{3-4}$$

using the power value p = 0.22 (Smulders et al., 2018). Across subjects, the chosen power value minimized Pearson's definition of *skewness* (i.e., distance between mean and median). Subsequently, the TFR was z-scored within each band across time for every block, to equalize variance across bands for better interpretation of the response model's β-weights (Figure 3-5B).

To get a general estimate of the dependency of power on SNR, we first contrasted the power spectra (time-frequency representations averaged across time) of the two unbalanced SNRs of –6 vs. +6 dB. We hypothesized to observe increased *alpha power* during –6 dB compared to +6 dB. On the subject level, within every channel-frequency bin, we contrasted the power distribution over 90 trials per SNR (–6 vs + dB; two-sided independent samples t-test). Across subjects, we bootstrapped the 95% confidence band of the mean t-values using 2000 iterations (Efron, 1979).

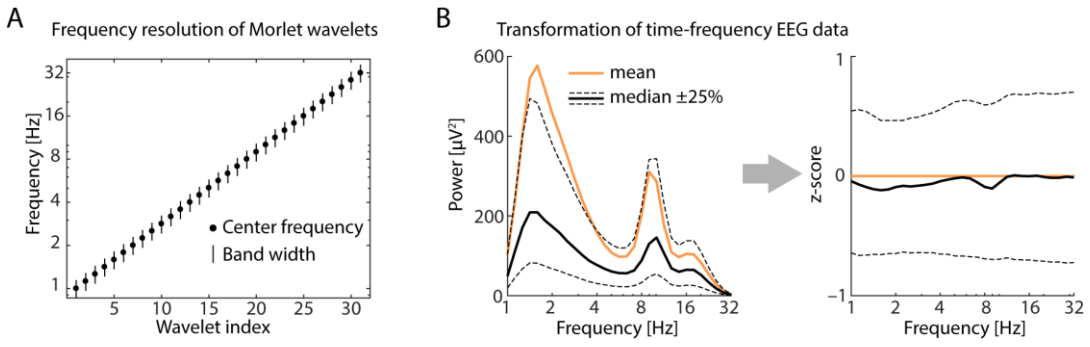

Figure 3-5: Time-frequency representation of EEG data. A) Logarithmically-spaced center frequencies (dots) and band widths (vertical lines) of Morlet wavelets. B) Band-wise distribution of power values across time before (left) and after transformation and z-scoring (right) of an exemplary subject.

3.2.3.2 SNR-related stimulus representations

The goal of this analysis was to estimate the relationship of the power time-courses (including induced neural oscillations) with the signal-to-noise ratio (SNR). Accordingly, we used the SNR as a predictor in a *forward model* (*SNR*, Figure 3-6, see section 3.2.3.3). The SNR is the ratio of sound intensity between the attended and the ignored talker and thus is an estimator for the current listening difficulty. It is important to note that this representation of the SNR is only describing the low-frequent modulation was created by scaling the magnitude of the talker signals. The representation of the SNR does not contain instantaneous fluctuations due to varying degrees of spectro-temporal overlap.

In an exploratory analysis, we used several stimulus representations derived from the SNR: First, we used the individual intensity of the talkers (*Levels*, Figure 3-6) as predictors, which reflects the hypothesis that *alpha power* is modulated by only one of the talkers. For example, we might observe only an increase in *alpha power* when the ignored talker gets louder, but not a decrease in *alpha power* when the attended talker gets louder. According to our initial hypothesis, we expected either a negative relationship between the level of the attended talker and *alpha power* or a positive relationship between the level of the ignored talker and *alpha power*.

Second, we used the first derivative of the talkers' intensity as predictors (*Change*, Figure 3-6), which is based on the hypothesis that *alpha power* is tracking only changes in listening conditions rather than the listening conditions in general.

Third, we only used the positive values of the first derivative of the talkers' intensity as predictors (*Increase*, Figure 3-6), which is based on the hypothesis that only increases in intensity of the (attended or ignored) talker lead to modulation of *alpha power*.

Fourth, we only used the negative values of the first derivative of the talkers' intensity as predictors (*Decrease*, Figure 3-6), which is based on the hypothesis that only decreases in intensity of the (attended or ignored) talker lead to modulation of *alpha power*.

Fifth, we used the envelope onsets as predictors, which also contain local modulation of the speech signals (*Env. onsets*, Figure 3-6). The underlying hypothesis was that a stronger decrease (i.e., desynchronization) of *alpha power* is tracking speech onsets of the attended talker, as it was found during visual covert attention (Jia et al., 2017).

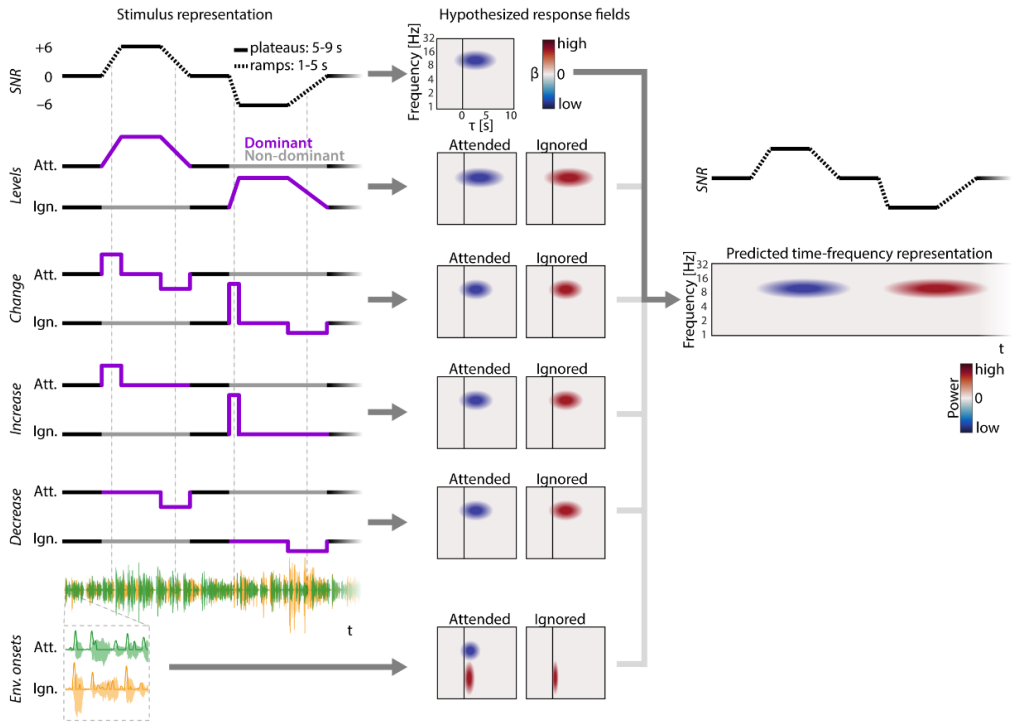

Figure 3-6: Representations of the signal-to-noise ratio used as regressors and hypothesized modulation of the EEG time-frequency representation. Left: SNR: signal-to-noise ratio. Levels: varying levels of the attended and ignored talker. Change: rate of change of the individual talker levels (fist derivative). Increase: positive change of the individual talker levels. Decrease: negative change of the individual talker levels. Envelope onsets: positive change of the individual talker envelopes. Middle: Hypothesized response fields. Right: Hypothesized time-frequency representation.

3.2.3.3 Estimation of time-frequency response fields

We trained *temporal response functions* (TRFs) per EEG channel and per frequency band, to estimate the (time-lagged) relationship between the stimulus representation (e.g., SNR) and the power time-courses within the frequency bands of the EEG signal. Per EEG channel, the training of the model resulted in a time-lag-by-frequency response field (Figure 3-6, middle). This response field shows the band-wise EEG power response to the stimulus representation (e.g., SNR). For example, a negative peak at a time lag τ of 3 seconds around 10 Hz would mean that power within the alpha band increases 3 seconds after the SNR decreases and/or the power in the alpha band decreases if the SNR increases. Except for the exact time lag of such a response, this example reflects our main hypothesis (Figure 3-6, *SNR*).

3.2.3.4 Prediction of time-frequency response

Analogous to the prediction of time-domain EEG signals (see section 2.5.1), we used the response fields to predict the EEG response. Here, instead of the broad-band time domain EEG signal, we predicted the power time course within each band. In order to obtain a measure for

neural selectivity, we predicted two EEG responses. The first EEG response was predicted with the actual stimulus representation as presented (true). The second EEG response was predicted with the opposite stimulus representation (false), which means that the SNR was flipped (for all other stimulus representations, *attended* was interchanged with *ignored*). If the EEG response predicted with the true stimulus representation yielded a higher Pearson-correlation coefficient with the measured EEG response than the false EEG response, the classification was counted as correct. The classification accuracy was expressed in a percentage and called *neural selectivity*. Note that here we used twelve five minute blocks per subject, such that the binomial chance level was 75%.

3.2.3.5 Reconstruction of SNR

The advantage of *forward models* is the interpretability of the response functions or response fields to multiple stimulus features (e.g., Level of the attended and the ignored talker). One disadvantage however, is that they only predict one neural response (here: EEG power time course at a certain frequency independent of all other frequencies). *Backward models* have the advantage of reconstructing one stimulus feature (e.g. SNR) based on multiple neural responses (e.g. all power time courses in the alpha range; see section 2.5.1). Given that all subjects may show enhanced *neural selectivity* somewhere in the alpha band but not at the identical frequency, we may have missed some general *alpha power* dynamics by using *forward models*.

Thus, exploratorily we trained *backward models* on the all frequency channels within the range of an octave, which roughly corresponds to conventional frequency bands (δ_{low} = 1–2 Hz; δ_{high} = 2–4 Hz; θ = 4–8 Hz; α = 8–16 Hz; β = 16–32 Hz). However, in contrast to conventional *backward models*, we trained the *backward model* not on all EEG channels but per EEG channel, such that an interpretation of the topography was still possible. Consequently, we could also compare band-wise *neural selectivity* to the *neural selectivity* obtained by the phase-locked responses.

3.2.4 Results

In this study, we investigated if the power in certain frequency bands of the EEG covaries with the signal-to-noise ratio (SNR) during selective attention to continuously presented and concurrent speech. We hypothesized that *alpha power* would be an indicator of the varying SNR. We trained *forward models* (i.e., response fields) in order to estimate the temporal dynamics of

EEG power relative to fluctuations of the SNR. We tested the predictive power of the response fields by detection of the focus of auditory attention.

3.2.4.1 No average power difference between SNRs

First, we contrasted the power spectra (time-frequency representations averaged across time) of the two extreme SNRs –6 vs. +6 dB. Within single subjects, we found channel-frequency bins where significance was reached (two-sided independent samples t-test, $p < 0.05$, uncorrected). The patterns of single-subject t-values are shown in Figure 3-7B. In general, the patterns show high variability across subjects. Contradictory to our hypothesis, only handful of subjects showed increased power in the alpha band during an SNR of –6 dB compared to +6 dB.

At the group level, we only found a few channel-frequency bins in the delta and theta range where power was found to be decreased during the SNR of –6 dB compared to +6 dB (bootstrapped mean of t-values, $p < 0.05$; Figure 3-7A). However, those differences did not survive correction for multiple comparisons. Consequently, a cluster-based permutation test including correction for multiple comparisons did not return any significant clusters.

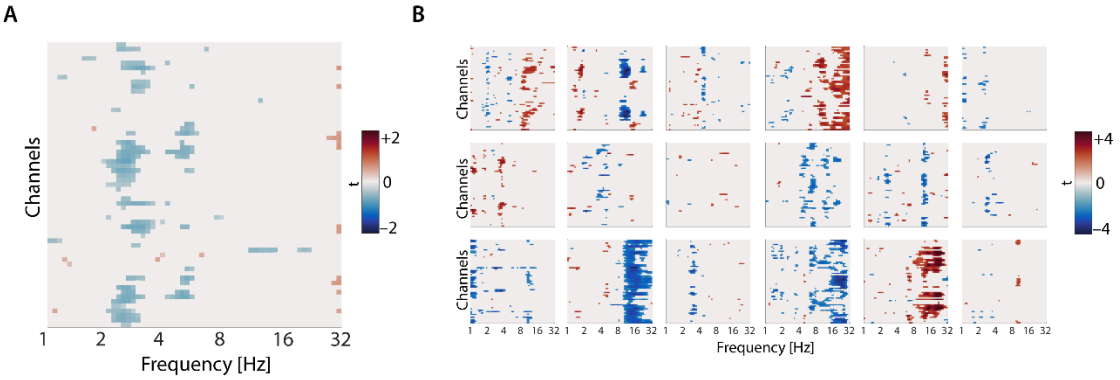

Figure 3-7: Difference of average power between SNR of –6 dB and +6 dB. A) T-maps (see B) averaged across subjects. Channel-frequency bins not significantly differing from zero were masked (one-sample t-test, dof: 17, $p<0.05$, uncorrected) B) Single subject t-maps of the difference of average power between –6 dB and +6 dB. Non-significant channel-frequency bins were masked (independent-samples t-test, dof: 178, $p<0.05$, uncorrected).

The initial analysis of the average power spectra during *plateaus* of –6 dB versus +6 dB did not confirm our hypothesis. However, with this initial analysis, the temporal dynamics of a potential *alpha power* modulation were neglected. For example, the neural response to a decreasing SNR (e.g., *ramp* to –6 dB) can be sluggish and thus smear into a subsequent *plateau* of 0 dB. Hence, we extracted band wise *temporal response function* in the following step.

3.2.4.2 No consistent modulation of alpha power by SNR

The temporal response fields reflect the time-lagged covariance between the SNR and the band-wise EEG power time courses. Positive peaks in the temporal response fields indicate that an increased SNR results in enhanced power, whereas negative peaks indicate that a decreased SNR results in enhanced power.

We hypothesized to find a negative relationship between the SNR and power in the alpha band. In general, the temporal response fields to the SNR show no unique peak, but rather a noisy, unspecific pattern (Figure 3-8A). In response to the SNR (first row), the most salient peak in the alpha band is a positive peak at around 10 seconds. Contradictory to our hypothesis, there is no evidence for an *alpha power* decrease when the SNR increases and vice versa.

Noteworthy, our random variation of the *plateau* length decreased predictability of the upcoming SNR, such that we would expect no peaks at time lags smaller than zero. However, in the extracted response fields, the appearance of peaks before zero indicates that mainly noise was fitted here rather than true relationships between the SNR and EEG power time courses (Figure 3-8A).

3.2.4.1 No significant modulation of alpha power by top-down selective processing

To estimate the effect of *top-down* attention on the neural response, we used the temporal response fields to predict the band-wise EEG power time courses and to detect the attended talker. Consequently, the percentage of trials the attended talker was correctly detected was termed *neural selectivity*. Enhanced *neural selectivity* within a certain frequency band indicates that power within this band is modulated by *top-down* attention.

We found slightly enhanced *neural selectivity* in the frequency range between 10 and 16 Hz (Figure 3-8B). However, the confidence band of the bootstrapped mean across subjects did not exceed the individual chance level based on a binomial distribution (12 trials; 75%). Furthermore, the topography did not show a focal pattern, which further indicates that the observed effects are due rather to common channel noise than a true relationship between the SNR and (for example parietal) *alpha power* modulation.

3.2.4.2 Exploratory analysis: Representations derived from the SNR did not reveal significant signatures of top-down selective processing

Exploratorily, we estimated the response fields to various derivates of the SNR (see section 3.2.3.2). First, we used the individual Level of the talkers as regressors (Figure 3-8C, first row). The response field to the attended talker shares some similarities with the earlier extracted response field to the SNR in the frequency range between 8 and 16 Hz. The inverse of the response field to the ignored talker shares some similarities with the earlier extracted response field to the SNR in the lower frequency range between 1 and 8 Hz. The channel-frequency pattern of *neural selectivity* turns out to be highly similar as well (Figure 3-8D, first row), which indicates that the individual power time courses do not selectively track the individual levels, for neither the attended nor the ignored talker.

Second, we estimated the response fields to the first derivative of the individual talker levels, which highlighted changes in the SNR (Figure 3-8, second row). The response fields suggest that a more negative deflection is following the attended talker around a time lag of 10 s compared to the ignored talker. This means that an increase of the level of the attended talker induces a stronger decrease in *alpha power* and/or a decrease of the level of the attended talker induces a stronger increase in *alpha power*. Yet again the attended–ignored response field difference shows multiple peaks and no consistent pattern. The alpha frequency range of slightly (but not significantly) enhanced *neural selectivity* is comparable to the results above, whereas its topography now shows right temporal channels with enhanced *neural selectivity* (Figure 3-8D, second row).

Third, we estimated the response fields to the positive parts (i.e., halfwave rectified) of the first derivative of the SNR, which highlighted increases in the talker levels (Figure 3-8C, third row). The response fields suggest that an increase of the level of both, the attended and ignored talker is followed by an increase in *alpha power*. Again, there is no salient difference between the response fields to the attended and ignored talker. The alpha frequency range of slightly (but not significantly) enhanced *neural selectivity* is comparable to the results above, whereas the topography shows no specific pattern (Figure 3-8, third row).

Fourth, we estimated the response fields to the negative parts (i.e., halfwave rectified) of the first derivative of the SNR, which highlighted decreases in talker levels (Figure 3-8C, fourth row). Note that we kept the negativity of the regressor after halfwave rectification, such that negative

weights in the response fields mean that a decrease of the talker level is followed by an increase of power. The response fields suggest that a decrease of the talker level is followed by an increase in *alpha power* independent of attention, we observed the same in the response to level increases. Again, there is no salient difference between the response fields to the attended and ignored talker. The alpha frequency range with slightly (but not significantly) enhanced *neural selectivity* is comparable to the results above, whereas the topography shows no specific pattern (Figure 3-8D, fourth row).

Fifth, we estimated the response fields to the envelope onsets of the attended and ignored talker (Figure 3-8C, fifth row). In contrast to the regressors above, the envelope onsets contain not only the low-frequent modulation related to the SNR, but also higher-frequent modulation (delta and theta) related to the onsets of syllables. The response field to the attended talker shows a decrease in *alpha power* immediately after the envelope onset. Slightly later, a decrease in power in the delta and theta band follows. At a time lag of approximately 7 seconds, an increase of *alpha power* is follows the envelope onset of the attended talker. The response field to the ignored talker shows a slightly different pattern with a more sustained *alpha power* decrease after the onset and a later increase of *alpha power* at a time lag of approximately 15 seconds. The attended–ignored contrast of the response fields suggests that the attended talker is followed immediately by decreased *alpha power* compared to the ignored talker. With later time lags, increased *alpha power* (and to some extend delta and theta power) can be observed in the response field to the attended compared to the ignored talker. However, this observed difference in the response fields does not lead to significantly enhanced *neural selectivity*, even if we again observed slightly enhanced *neural selectivity* in the alpha range (Figure 3-8D).

In sum, the exploratory analysis did not result in more precise predictions of the attention-and SNR-dependent modulation of neural oscillatory power. However, all predictions indicated slightly enhanced *neural selectivity* in the alpha frequency range.

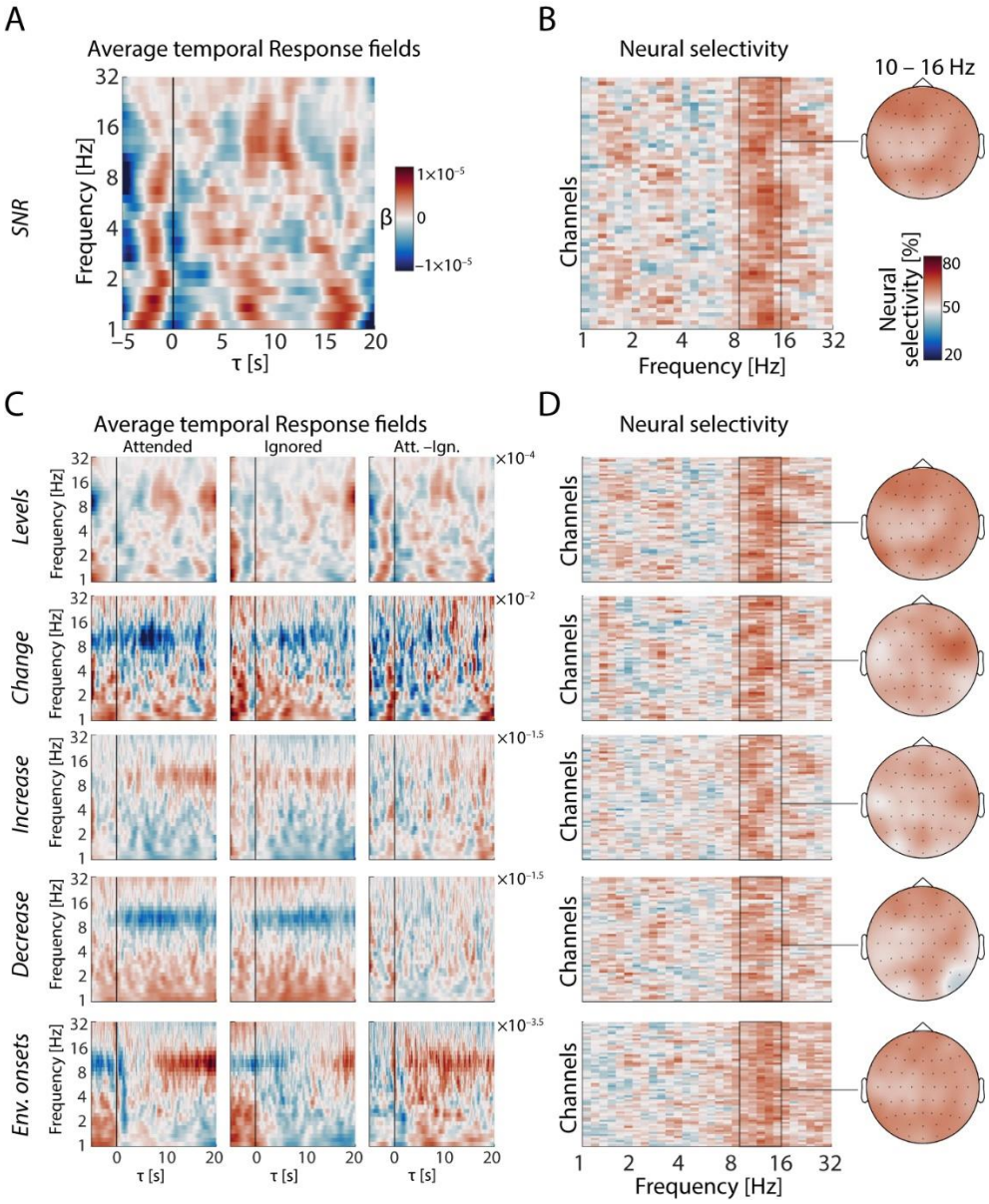

Figure 3-8: Time-frequency response fields and *neural selectivity*. A) Response field to the SNR. B) *Neural selectivity*. The lower confidence bound after bootstrapping the mean across subjects did not reveal any significant channel-frequency bins (2000 iterations, binomial chance: 75%, $p < 0.05$, uncorrected). Topographic map shows *neural selectivity* averaged across 10–16 Hz. C) Response fields to derivates of the SNR (see methods). D) *Neural selectivity*. The lower confidence bound after bootstrapping the mean across subjects did not reveal any significant channel-frequency bins (2000 iterations, binomial chance: 75%, $p < 0.05$, uncorrected). Topographic maps show *neural selectivity* averaged across 10–16 Hz.

3.2.4.3 Exploratory analysis: Reconstruction of the SNR within the alpha band did not reveal significantly enhanced neural selectivity

In order to use the full predictive (i.e., reconstructive) power of a whole frequency range (e.g., alpha band) and allow for individual best-frequency-differences within said range, we trained *backward models* by using multiple frequency channels of the time-frequency representation as regressors to reconstruct the SNR. We split up the time-frequency representation into five bands

with each band spanning an octave (δ_{low} = 1–2 Hz; δ_{high} = 2–4 Hz; θ = 4–8 Hz; α = 8–16 Hz; β = 16–32 Hz).

Neural selectivity was slightly enhanced in the alpha and beta band compared to the lower frequency bands (Figure 3-9, left). However, in none of the frequency bands, the lower confidence bound of the bootstrapped mean across subjects exceeded the empirical chance level of 75% at any EEG channel (2000 iterations, $p<0.05$). Again, the topographic maps show an unspecific, non-focal pattern.

We compared the *neural selectivity* estimated in the alpha band to *neural selectivity* estimated in the phase-locked responses (Figure 3-9). Strikingly, the phase-locked responses yielded *neural selectivity* above the empirical chance level at almost all EEG channels. In each subject, at least 9 of 62 EEG channels exceeded the empirical chance level (mean: 46.94 channels, SD: 18). Divergently, in 17 of 18 subjects, *neural selectivity* of the alpha band exceeded the empirical chance level at only two EEG channels (mean: 18, SD: 16.59). The topographical inconsistency of *neural selectivity* in the alpha band does not allow a selection of channels of interest for further comparison. Averaged across all EEG channels, *neural selectivity* of the phase-locked responses exceeded the empirical chance level in 13 of 18 subjects, whereas *neural selectivity* of the alpha band exceeded the empirical chance level in only 3 subjects (Figure 3-9). Pearson-correlation of the *neural selectivity* averaged across all channels between phase-locked responses and *alpha power* was marginally significant (Pearson's $r = 0.41$, $p = 0.088$), which indicates that subjects with enhanced *neural selectivity* of the phase-locked responses also show enhanced *neural selectivity* in the alpha band. Phase-locked responses yielded significantly enhanced *neural selectivity* compared to *alpha power* in all but the occipital EEG channels (two-sided paired samples t-test, dof: 17, $p < 0.05$). In sum, this comparison highlights that *neural selectivity* of the phase-locked responses is an order of magnitude higher than *neural selectivity* of the alpha band.

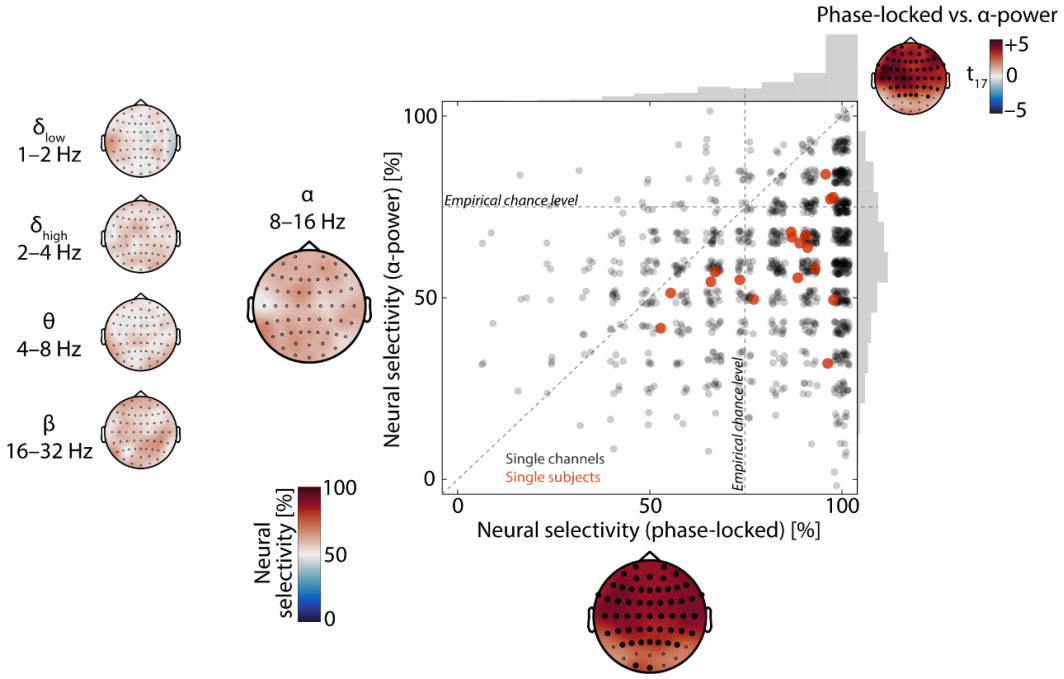

Figure 3-9: *Neural selectivity* obtained by backward reconstruction of SNR time courses within frequency bands and comparison to *neural selectivity* obtained by prediction of phase-locked neural responses. Topographic maps show channel-wise *neural selectivity*. Scatter plot compares *neural selectivity* obtained by prediction of phase-locked responses (x-axis) versus alpha-power-based reconstruction of the SNR (y-axis) at EEG channels of all subjects (grey) and average across all EEG channels per subject (red). For better illustration of density, underlying values were jittered by uniformly distributed values of ±2%. Histograms show distribution of *neural selectivity* across all single-subject EEG channels. Highlighted channels of the topographic maps exceeded the binomial chance level of 75% (lower bound of the confidence band obtained by bootstrapping the mean across subjects), which was only found in the phase-locked response. Upper-right topographic map shows t-values (two-sided paired samples t-test, dof: 17) of contrasted *neural tracking* between the phase-locked response and *alpha power*. Highlighted channels show a significant difference ($p < 0.05$).

3.2.5 Discussion

Here we investigated whether the modulation of neural oscillatory power depends on the signal-to-noise ratio (SNR). Based on previous studies, we hypothesized that *alpha power* is inversely covarying with the SNR, such that a positive SNR results in decreased *alpha power* and a negative SNR results in increased *alpha power*, reflecting the demand for *top-down* attentional control. Since we had no precise hypothesis about the temporal dynamics of *alpha power* relative to the SNR, we estimated *forward models* (time-frequency response fields), which reflect the time-lagged relationship between the SNR and the EEG time-frequency representation (i.e., narrow-band power time courses). Our analysis revealed no clear evidence, neither on the single subject nor on the group level, for *alpha power* modulation in the hypothesized direction. The estimated time-frequency response fields did not show consistent patterns that allowed inference on the neural oscillatory dynamics under continuously varying listening conditions (i.e., SNR). However, we observed slightly, but not significantly enhanced *neural selectivity* in the *alpha power*

range with a non-focal, unspecific distribution across the scalp. Exploratory analysis based on various regressors did not reveal any conclusive results either. An EEG-channel-wise comparison between *neural selectivity* of the alpha band versus phase-locked responses demonstrated that *top-down* neural selective processing of the concurrent speech is much more prominent in the phase-locked responses.

3.2.5.1 Alpha power modulation in trial-based but not continuous designs

Our hypothesis was based on recent, trial-based studies which associated more adverse listening conditions with enhanced *alpha power* (Obleser et al., 2012, Obleser und Weisz, 2012; Wöstmann et al., 2017b; McMahon, 2016). However, during our continuously presented speech, we did not find clear evidence for such a modulation indicating the demand for *top-down* attentional control. In the following, we will discuss possible reasons for such an indistinct finding.

One possible cause might be the general structure of our design. While earlier studies used trial lengths of a couple of seconds, we presented the speech signals in blocks of five minutes. In trial-based designs, increased *alpha power* can be usually observed in the beginning compared to the end of each trial (Wöstmann et al., 2015; Wöstmann et al., 2016), which suggests that *alpha power* modulation reflects anticipation of the upcoming neural processing or, in other words, adaptation of the neural filters, rather than the neural processing itself. In the current design, neural filters might have been adapted quickly in the beginning of each block, such that *alpha power* modulation was absent during the rest of the block. However, we also failed to extract attention-dependent *alpha power* modulation during changes of the SNR, which would also have induced a re-adaptation of neural filters. This might be explained by the fact that *alpha power* occurs in bursts, which might not occur time-locked to changes in the SNR, but rather at any time during a *plateau*. However, we found no consistent differences between the across-time-averaged time-frequency representations under −6 versus +6 dB, which should capture such non-time-locked relationships.

Another explanation might be the stronger engagement of subjects into a trial-based task compared to a continuous task. Increased *alpha power* has been related to increased working memory load (Leiberg et al., 2006; Karraschet et al., 2004; Obleser et al., 2012). It has been argued that increased *alpha power* during presentation of degraded speech reflect enhanced lexical

memory access (Obleser and Weisz, 2012). In the trial-based designs, subjects usually repeat some words such as digits, which leads to repeated engagement into the task. It was also shown that *alpha power* lateralization follows the temporal structure of the stimuli (Wöstmann, 2016). In contrast, during our task, subjects only had to answer four questions in the end of each block, which may not have challenged them as much as in the trial-based designs. The risk of missing a few words and not being able to answer the subsequent question might have been too low to force participants to invest as much of their cognitive resources. This point will be further discussed below (see section 0).

The slight enhancement of *neural selectivity* found in the alpha band might indicate that there exists a relationship between the SNR and *alpha power*, which did not reach significance because of an underpowered study design. As noted above, the effects found in previous studies are strongest during anticipation of an upcoming stimulus and tend to decay during towards the end of the trial. Hence, continuous presentation might lead to a weaker modulation of *alpha power* which only reaches significance on the group level based on a greater number of subjects. Henceforth, we can conclude that *alpha power* is not as informative about a listener's focus of auditory attention as the phase-locked responses to speech.

3.2.5.2 Conclusion

We showed that there exists no consistent pattern of *alpha power* modulation that allows for the conclusion that enhanced *alpha power* indicates a stronger demand for *top-down* attentional control. Thus, the link of *alpha power* modulation to the concept of listening effort is not supported by this study. However, a slight enhancement of *neural selectivity* was found in the alpha band, which may be a hint that a small effect size led to an underpowered design. In contrast to the phase-locked neural responses, *alpha power* is not as informative about the listener's focus of attention, which has implications on the application in neurally steered hearing aids.

3.3 Study 3: No *alpha power* lateralization induced by continuously moving talkers

3.3.1 Abstract

Alpha power lateralization has been observed during spatial selective attention tasks in the visual, somatosensory and auditory modality and has been associated with induced, *top-down* neural mechanisms. Here we first asked whether *alpha power* lateralization is indicative of a listener's spatial focus of auditory selective attention. We then asked if *alpha power* lateralization interacts with varying signal-to-noise ratio (SNR). To this end, we recorded the electroencephalogram (EEG) of 25 subjects who listened to one of two simultaneously presented stories, while the talkers stochastically moved on the frontal azimuth between −90 and +90 degrees and the SNR between the attended (signal) and the ignored talker (noise) varied between −6 and +6 dB. First, we hypothesized that *alpha power* lateralization can be observed when the talkers are located on opposite positions. Second, we hypothesized that *alpha power* lateralization should be enhanced when the SNR is worse due to the stronger demand for *top-down* attentional control. To test our hypothesis, we forward-modelled the time-frequency representation of the EEG signal based on the location difference and the SNR. We did not find any indication for *alpha power* lateralization nor the interaction with SNR.

3.3.2 Introduction

Alpha power lateralization has been observed during spatial selective attention tasks in the visual (Worden et al., 2000), auditory (Kerlin et al 2010) and tactile modality (Haegens et al., 2011). Due to its non-phase-locked nature, *alpha power* lateralization is associated with *top-down* attentional, induced neural mechanisms. In particular the observation of lateralized parietal *alpha power* during dichotic listening tasks (Kerlin et al., 2010; Wöstmann et al. 2016) supported the hypothesis that the functional role of *alpha power* is *top-down* inhibition of brain areas that are *bottom-up* captured by task-irrelevant sensory input (Klimesch et al., 2007).

Our earlier study could not confirm our hypothesis that *alpha power* is indicative of the current demand for *top-down* attentional control during the presentation of continuous speech, which we manipulated by the signal-to-noise ratio (SNR; see section 3.2). One possible reason we have not observed such an *alpha power* modulation might be the diotic presentation of the talkers. The lack

of distinct predictable spatial (e.g., Wöstmann et al., 2016) and temporal cues (e.g. Wöstmann et al., 2018) might have hindered *alpha power* to operate as an inhibitory mechanism suppressing neural pathways occupied by the processing of the ignored talker.

Consequently, we adopted our earlier paradigm and added virtual acoustic scene, where not only the SNR between the talkers but also the location of the talkers stochastically varied. First, we hypothesized to find *top-down* modulated *alpha power* lateralization by means of enhanced *alpha power* contralateral to the ignored talker and/or decreased *alpha power* contralateral to the attended talker. Second, we hypothesized that *alpha power* lateralization interacts with SNR due to the varying demand for attentional control, such that we would observe stronger *alpha power* lateralization during and SNR of –6 dB compared to +6 dB.

3.3.3 Methods

3.3.3.1 Participants

We recruited 25 native speakers of German within the age range of 18 to 31 (mean: 22.7 years, 15 female). All reported normal hearing and no histories of neurological disorders. We recorded each subject's pure-tone audiogram (PTA) in order to verify normal hearing. None of the subjects showed a PTA of more than 20 dB within the frequency range of 125 to 4000 Hz. All participants gave informed consent and received payment of 8 €/hour. The data of one participant had to be discarded due to technical issues during the experiment.

3.3.3.2 Stimuli

Analogous to our earlier study where we stochastically varied only the signal-to-noise ratio (SNR) during continuous presentation of audiobooks (Fiedler at al. 2019; see section 3.1), here we additionally varied the location of the talkers by simulating continuous movements (Figure 3-10). The talkers independently moved along the frontal azimuth between two most lateral locations (–90° and +90°, 0° elevation), resulting in four principle constellations (Figure 3-10, left). In combination with the three different SNRs (–6, 0 and +6 dB), twelve different conditions existed.

We individually randomized the continuous time course through the twelve conditions for every subject (Figure 3-10, middle). Analogous to our earlier study, the condition was kept constant during a *plateau*. A *plateau* lasted between five and nine seconds (uniformly distributed

in discrete steps of one second) and the transitions between *plateaus* (further called *ramps*) took between one and five seconds (uniformly distributed in discrete steps of one second). During a *ramp*, either the SNR or the location was changed. Within the session of each subject, every possible *ramp* occurred the same amount of times.

Analogous to our earlier study (Fiedler at al. 2019; see section 3.1), we selected two audiobooks read by native German speakers, one female (Aude Le Corff, 'Das zweite Leben des Monsieur Moustier, read by Claudia Drews) and one male (Peter Wohlleben, 'Das geheime Leben der Bäume', read by Roman Roth). Sequences of silence longer than 500 ms were truncated to 500 ms to avoid long periods of silence (O'Sullivan et al., 2014). The first hour of each audiobook was selected for further preparation. The first 30 minutes of each audiobook served as the to-be-attended and the rest served as the to-be-ignored speech, such that all subjects could attend both stories from the beginning and attended (and ignored) both the female and the male voice the same amount of time. The one-hour audiobooks were split up in 12 blocks of approximately five minutes each.

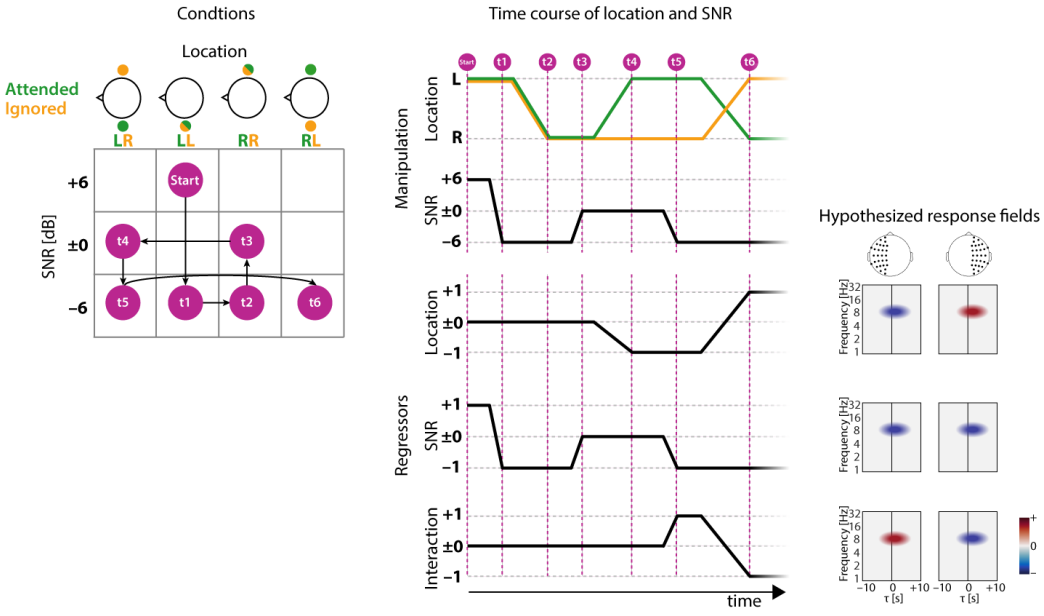

Figure 3-10: Exemplary sequence of SNR and location. Left: three-by-four design of three levels of the signal-to-noise ratio (SNR; −6, 0 and +6 dB) and four spatial constellations of the talkers (LR, LL, RR, RL). Middle: The continuous time course through the conditions was randomized for every subject individually. Based on the time course, regressors were defined which represented the SNR, the location and their interaction. Right: Hypothesized hemispherical time-frequency response fields estimated based on the regressors.

3.3.3.3 Spatial presentation via head-related transfer functions

The spatial movement was realized by the creation of virtual sound sources using *head-related transfer functions* (HRTFs). We used HRTFs from a database recorded in the ear canal of 97

subjects (In-The-Ear HRTFs, Institut für Schallforschung, Vienna, Austria). The database contains binaural impulse responses with a spatial resolution of 2.5° in the frontal azimuth.

Theoretically, the presentation of a sound waveform convolved with the binaural impulse responses of a certain location should be correctly localized by a subject. However, since head shapes largely differ between individuals, one HRTF might create a better spatial impression than another. Typical artifacts of non-individualized HRTFs are front-back confusions and elevation of the virtual sound source. Hence, the best-fitting HRTF should be selected for each subject. To reduce the set of HRTFs to choose from, we first calculated the Pearson-correlation coefficients between the impulse responses of two HRTFs for each location and ear. Subsequently, correlation coefficients were averaged across all locations. This was done for all possible pairs of all HRTFs. We used multi-dimensional scaling (Kruskal, 1964) of the inverted magnitude of the Pearson correlation coefficients to obtain a two-dimensional representation of the dissimilarity between all 97 HRTFs. We manually selected five most different HRTFs for further evaluation by the subjects (see below; selected HTRF index: 4, 66, 92, 123, 162).

To make out the best-fitting HRTF for every subject, in advance of the main experiment, subjects were presented with ten-second probes of audiobooks with a female talker (Elke Heidenreich, 'Nero Corleone kehrt zurück', read by Elke Heidenreich) or a male talker (Yuval Noah Harari, 'Eine kurze Geschichte der Menschheit', read by Jürgen Holdorf) moving from the left to the right side (±90°) or vice versa along the frontal azimuth. In advance, subjects were informed about typical artifacts such as front-back confusions and elevation of the virtual sound source. After each presentation, subjects were asked to rate the spatial quality on a scale from 1 to 4. In total, 20 probes were presented (5 HRTFs × 2 directions × 2 talkers). The HRTF with the highest average rating was selected for the subject. All subjects reported that they heard at least one presentation where the talker moved in the frontal hemifield along the azimuth. The distribution of the selected HRTFs (9;4;6;2;3) indicates that subjects preferred different HRTFs.

To create a percept of continuous movement, we first created the binaural sound wave forms of all discrete locations from −90° to +90° in steps of 2.5°. Based on the randomized trajectories of the talkers resulting from the *ramps* between locations, we linearly interpolated the sound wave forms between the two discrete neighbored locations.

3.3.3.4 Task

The twelve blocks were presented such that subjects were instructed to attend to the female or to the male talker in an alternating fashion. Every other block, the stories picked up at the point at which ended two blocks before. Whether the first instruction was "attend female" or "attend male" was counterbalanced across subjects. After instruction before each block (i.e. attend to female or attend to male), subjects were asked to start the stimulus presentation by a button press, which enabled the participants to take a break between blocks. During listening, subjects were asked to fixate at a cross presented on the screen in order to reduce eye movement. After each block, participants were asked to answer four multiple-choice questions concerning the content of the to-be-attended audiobook.

3.3.3.5 Data acquisition and preprocessing

EEG was recorded with 64 electrodes *Acticap* (*Easycap*, Herrsching, Germany) connected to an *ActiChamp* (*Brain Products*, Gilching, Germany) amplifier. EEG signals were recorded with the software *BrainVision Recorder* (*Brain Products*) at a sampling rate of 2.5 kHz. Impedances were kept below 10 kΩ. Electrode TP9 (left mastoid) served as reference during recording.

The EEG data were pre-processed in *MATLAB 2017a* (*The MathWorks, Inc.*, Natick, Massachusetts, United States) using both the *Fieldtrip*-toolbox (version: 20170321; Oostenveld et al., 2011) and custom written code. The EEG data were re-referenced to the average of the electrodes TP9 and TP10 (left and right mastoids) and resampled to fs = 125 Hz. The continuous EEG data were highpass-filtered at fc = 1 Hz and lowpass-filtered at fc = 40 Hz (two-pass Hamming window FIR, filter order: 3fs/fc).

From the continuous EEG data, we extracted the parts during which the twelve blocks of audiobooks were presented (see above). For every subject, we applied independent component analysis (ICA; Makeig et al., 2004) on the concatenated data of the twelve blocks and manually rejected components that were clearly related to eye movements, eye blinks, muscle artifacts, heartbeat as well as single-channel noise.

3.3.3.6 EEG time-frequency representation

The time-frequency representation (TFR) of the pre-processed EEG data was estimated analogous to our previous study (see section 3.2), including the transformation to correct for the

skew of power distribution (Smulders et al., 2018) and subsequent z-scoring. Only for an exploratory analysis of the lateralization index (see below), did we keep the absolute power values.

To get a general estimate of the dependency of *alpha power* on SNR as well as location and the interaction of the two latter, we first contrasted the power spectra (time-frequency representations averaged across time) of the two unbalanced SNRs of −6 vs. +6 dB and the two separated locations (*LR* vs. *RL*). On the subject level, the variance of every channel-frequency bin was analyzed (Two-way ANOVA, two-by-two design). In order to test whether a resulting F-statistic is significant or simply a random observation, we compared it to a distribution of F-values obtained from surrogate data (2000 random permutations of condition labels). If less than 5% of the surrogate F-values exceeded the F-statistic obtained from the true condition labels, the channel-frequency bin was assumed to show a significant main effect or interaction, respectively.

3.3.3.1 Prediction of time-frequency response

Analogous to our earlier study (see sections 3.2), we estimated time-frequency response fields to predict the EEG response, but here we predicted the EEG power time course based on the SNR as well as the location of the talkers. To this end, we first defined the regressors. The representation of the SNR was scaled between −1 and +1. The location regressor was set as the difference between the locations of the talkers scaled between 0 (left) and +1 (right), such that −1 represented *LR* (attended talker on the left, ignored talker on the right) and +1 represented *RL* (attended talker on the right, ignored talker on the left). Consequently, during the *plateaus RR* and *LL*, the location regressor was zero. A third regressor expressed the hypothesized interaction between location and SNR. We multiplied the location regressor with the SNR regressor. For example, during an *LR plateau* under an SNR of −6 dB, the regressor was −1. If the SNR changed to +6 dB, the interaction regressor changed to +1. Now if the location changes to *RL*, the interaction regressor became −1.

To obtain a measure for neural selective processing, we predicted two EEG responses. The first EEG response was predicted with the actual stimulus representation as presented (true). The second EEG response as predicted with the opposite stimulus representation (false), which means that all regressors were flipped between *attended* and *ignored*. If the EEG response predicted with the true stimulus representation yielded higher Pearson correlation with the measured EEG response than the false EEG response, the classification was correct. The classification accuracy was expressed in percent (number of trials correct) and called *neural selectivity*.

3.3.3.2 Lateralization Index

Exploratorily, we calculated the ongoing lateralization index (LI). The LI expresses the time-resolved hemispheric imbalance of EEG power. In the alpha band, the LI was found to be indicative of the listeners' focus of auditory attention during dichotic listening (Wöstmann et al., 2016). For every frequency, we calculated

$$LI = \frac{left - right}{left + right} \quad \text{3-5}$$

where *left* (*right*) is the average power at the left (right) occipito-parietal channels. We forward-modelled and predicted the LI based on the three regressors locations, SNR and their interaction (see above). According to our hypothesis, mainly the location regressor and the interaction regressor should be predictive of the lateralization index. The predictability was again expressed in the percentage of correctly classified trials and called *neural selectivity*.

3.3.4 Results

Here we investigated whether *alpha power* lateralization can be observed during continuous and independent movement of an attended (signal) and an ignored (noise) talker and whether *alpha power* lateralization interacts with the signal-to-noise ratio (SNR). To this end, we presented two talkers simultaneously which stochastically moved along the frontal azimuth between −90° and +90°. The SNR was stochastically varied between −6 and +6 dB.

3.3.4.1 No average power difference between lateralized locations or SNRs

First, we investigated how across-time averaged power is modulated by the location and the SNR as well as whether there exists an interaction between location and SNR (Two-way ANOVA; see methods). On the single subject level, we found significant channel-frequency bins for the main effect of SNR, location and their interaction (Figure 3-11, $p < 0.05$, uncorrected). However, on the group level, we found no channel-frequency bin where any of the main effects or the interaction exceeded the critical F-value. This indicates that average power is neither modulated by the SNR nor the spatial focus of attention.

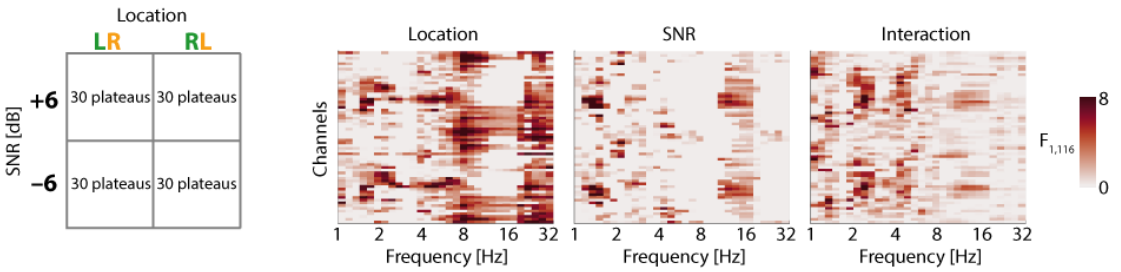

Figure 3-11: Main effect of location and SNR and their interaction. Left: The power within each *plateau* was averaged across time and the four acoustically equivalent conditions were contrasted in a two-by-two ANOVA. Right: Channel-frequency F-maps obtained from an exemplary subject. Values not exceeding the critical F-value (2000 permutations, 95%, uncorrected) were masked. On the group level, the F-maps averaged across subjects did not exceed the critical F-value at any channel-frequency bin.

3.3.4.2 The temporal dynamics of EEG alpha power do not indicate listeners' focus of attention

We *forward-modelled* and predicted the EEG time-frequency representation based on the regressors SNR, location and their interaction. We hypothesized that the temporal dynamics of *alpha power* are predictive of a listener's focus of attention. The predictability was expressed as the percentage of correctly classified blocks and called *neural selectivity*.

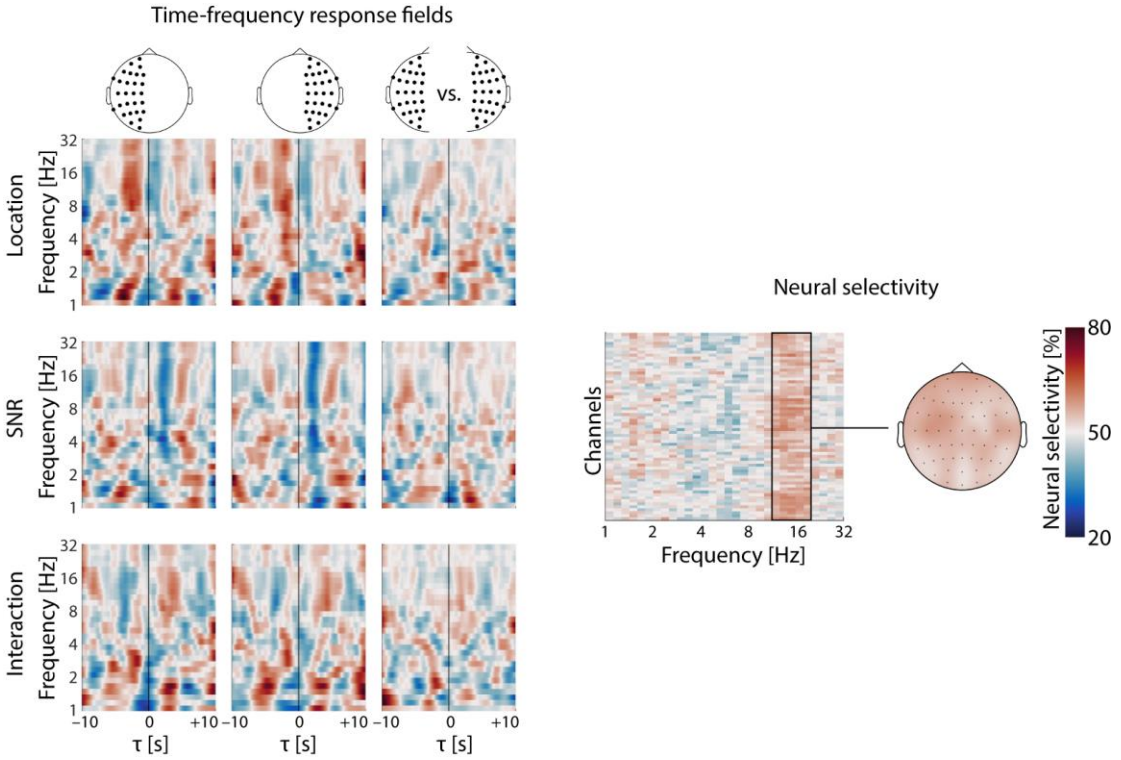

Figure 3-12: Time-frequency response fields and *neural selectivity*. Left: Hemispherical time-frequency response fields to location, signal-to-noise ratio (SNR) and their interaction. Right: *Neural selectivity* obtained at single frequencies and single EEG channels.

The time-frequency response fields express the spatio-temporal relationship between the regressors and the time-frequency representation of the EEG signal (Figure 3-12). We separately inspected the response fields for left and right EEG channels, since the β-weights in the response field to the location regressor should appear anti-polar if *alpha power* lateralization is modulated by the spatial focus of attention (see methods). However, the contrast between the two hemispheres did not reveal any salient peaks in the *alpha power* range.

3.3.4.3 *Hemispherical alpha power lateralization is not predictive of the listener's focus of attention*

In an exploratory analysis, we forward-modelled and predicted the time-resolved parietal lateralization of EEG power (Lateralization Index; see methods). In contrast to our hypothesis, we did not find enhanced *neural selectivity* in the *alpha power* range (Figure 3-13). *Neural selectivity* averaged across subjects remained around chance (50%) across the whole frequency range. Individual subjects show a random pattern, which rarely exceeded the binomial chance level of 75%.

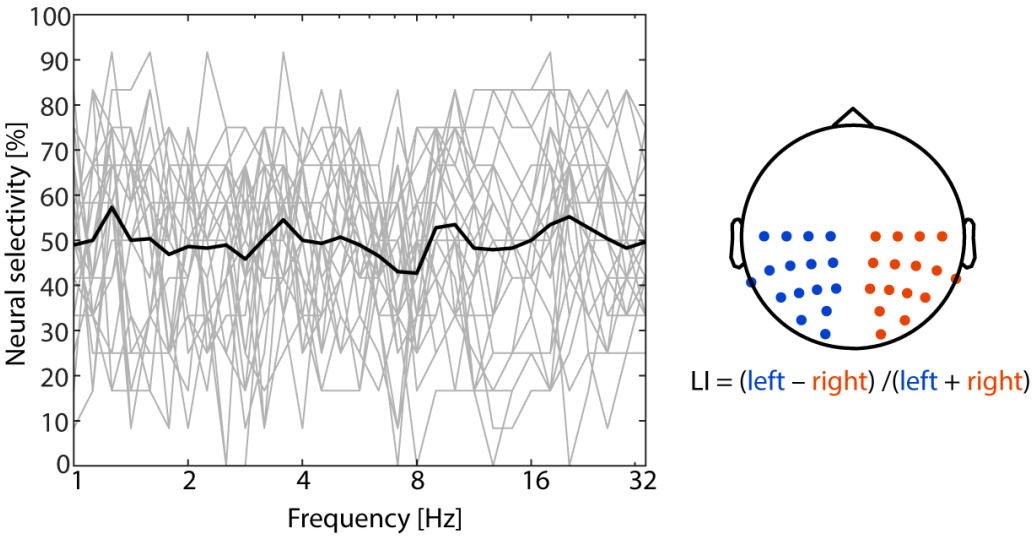

Figure 3-13: *Neural selectivity* obtained by the prediction of the hemispherical lateralization of EEG power. Left: *Neural selectivity* obtained from the prediction of the lateralization index (LI) at single frequencies. Right: The calculation of the LI was based on the average across left and right occipito-parietal channels, respectively.

3.3.5 Discussion

Here we investigated if *alpha power* lateralization can be observed during the presentation of continuous speech from concurrent talkers moving stochastically along the frontal azimuth in a virtual acoustic space. We hypothesized that *alpha power* lateralization is indicative of the

listener's spatial focus of attention and that there exists an interaction with the orthogonally varied signal-to-noise ratio (SNR) between the attended (signal) and the ignored (noise) talker. We did not find any indication that the lateralization of *alpha power* can be predicted based on the spatial focus of attention. Furthermore, there was neither an indication for the overall modulation of *alpha power* by the SNR nor the interaction of the SNR and the location of the talkers.

3.3.5.1 No lateralization of alpha power during continuous presentation of an auditory scene

To our knowledge, this is the is the first study that investigates the role of *alpha power* lateralization during selective listening to concurrent, continuous speech of randomly moving talkers. The lack of attention-driven *alpha power* lateralization during continuous listening leads to open questions about the involvement of *alpha power* into attentional filtering as well as to the discussion of our experimental design.

Since *alpha power* lateralization was shown to be driven by the spatial focus of attention in trial-based designs such as dichotic listening to digit streams (Wöstmann et al., 2016), we must conclude that there may exist different neural strategies to overcome such a listening task. The proposed inhibitory functionality of *alpha power* might not be the neural strategy that is involved here.

One possible reason might be that here listeners were not forced to invest as much cognitive effort as in a trial-based design. In our design, subject only had to answer four questions after five minutes of listening, whereas in trial-based designs, subjects are asked much more frequently to respond. This might lead to frequent (re-)engagement into the task. In a recent study, it was shown that listening to clear, continuously moving pink noise with embedded targets leads to *alpha power* lateralization (Bednar and Lalor, 2018). This is somewhat surprising, since no distracting stimulus was presented. However, it strengthens the argument that subjects must engage into a more behaviorally fine-grained task, such that *alpha power* is involved in spatial filtering.

The current findings will be further discussed below within the whole framework attentional modulation of phase-locked neural response and *alpha power* (see section 5.3).

3.3.5.2 Conclusion

We investigated the role of *alpha power* during auditory selective attention to one of two randomly moving talkers, while the SNR between the attended talker (signal) and the ignored talker (noise) was varied orthogonally. We did not find clear evidence that *alpha power* is key to selective processing of continuous, dynamic auditory scenes. This has implications on the neural implementation of attentional filters and neurally steered hearing aids.

4 In-ear EEG captures signatures of auditory attention

In this chapter, we will show that in-ear EEG is feasible to extract neural responses to continuous stimuli and that those neural responses are informative of the listener's attentional focus.

Study 4 will show that single-channel in-ear EEG captures (spectrally resolved) neural responses. Study 5 & 6 will show that single-channel in-ear EEG captures neural responses to attended and ignored stimuli and that those neural responses indicate which out of two stimulus streams is attended.

4.1 Study 4: In-ear EEG captures spectrally resolved responses to natural stimuli[2]

4.1.1 Abstract

Spectro-temporal response functions (STRFs) reflect the average unfolding of a neural response to the spectro-temporal modulation of auditory stimuli. Given that the frequency-dependent impact of a sensorineural hearing loss modulates STRFs, the latter might be used to estimate a hearing loss and to fit a hearing aid. Furthermore, STRFs could be estimated in real-time and a hearing aid might be adapted. The goal of this study was the estimation of neural responses to natural auditory stimuli (i.e., rich of spectro-temporal modulation) via in-ear EEG. The EEG electrode configuration should consist only of a reduced set, to test if such a configuration could be attached to a hearing aid. A random sequence of sounds was presented to six subjects wearing individually fitted in-ear EEG electrodes. Subjects were asked to detect repeating sounds (one-back task). STRFs were trained by forward-modelling the frequency-dependent neural response via ridge regression. The EEG signals were predicted, and the measures *neural tracking* and *prediction accuracy* were estimated. The STRFs showed a succession of P1-N1-P2 response components, with stronger magnitudes in the response to lower frequencies. Compared to P1 and N1, the P2 component was most prominent. At least three out of six in-ear EEG channels in all but one subject showed *prediction accuracy* above chance. A strong relationship between the

[2] *The underlying data (Experiment 2) was conducted and pre-analyzed by Stephan Müller, Daniel Bank and Raphaela C. Wurzer (Fachhochschule Lübeck, course: Hörakustik, year of matriculation: 2014).*

strength of *neural tracking* and the *prediction accuracy* was found across subjects. In sum, the results show that stimulus-frequency resolved neural responses to sound can be estimated by single-channel configurations including in-ear-EEG electrodes.

4.1.2 Introduction

Note: The following two paragraphs are adopted from Wöstmann et al. (2017a) and were primarily drafted by Lorenz Fiedler.

Supported by increasing computational resources and efficient algorithms, the growing field of brain-computer interfaces has moved M/EEG towards real-time applications. Recent studies showed an attention-dependent *neural tracking* of speech (Ding and Simon, 2012; O'Sullivan et al., 2014, Mirkovic et al., 2015). This has encouraged the development of brain-controlled hearing aids (Lunner and Gustafsson, 2016). Thus far, hearing aids are amplifying the incoming sound regardless of the listener's intent. A detection of the attended talker based on the listener's neural response might be used to steer directional microphones or noise suppression of a hearing aid. Moreover, researchers make effort on the development of portable EEG devices, including the shrinking and hiding of sensor units to achieve higher convenience and thus a lower barrier of acceptance for users (Debener et al., 2015).

There are several challenges in these real-world applications to be addressed in the near future: (1) low SNR caused by a lack of shielding, movement artefacts, and unreliable electrode connectivity; (2) low number of channels hinders localization of contributing cortical regions; and (3) real-world hearing scenarios are manifold and change unpredictably and thus might not be comprehensively addressed by abstract scenarios presented in the lab.

While EEG detection of auditory attention is mainly based on the neural phase locking to the broad-band temporal envelope, the stimulus-frequency dependent neural response can also be extracted from M/EEG in the form of *spectro-temporal response functions* (STRFs; Ding and Simon, 2012). STRFs reflect the average neural response to sound-intensity fluctuations resolved in certain frequency bands. As in ERPs and TRFs, the components of the STRFs can be interpreted as representatives for processing stages along the auditory pathway. Hence, in-ear EEG might also serve as a diagnostic tool for conductive and sensorineural hearing loss.

Here we asked whether STRFs can be extracted from in-ear EEG and if increased spectral resolution of the stimulus leads to prediction accuracy above chance. We show that STRFs can be extracted from in-ear EEG. The STRFs show frequency-dependent modulation. However, the increased spectral resolution did not lead to a clear enhancement of *prediction accuracy*.

4.1.3 Methods

4.1.3.1 Participants

Six normal-hearing participants were enrolled in this study. In advance, they underwent a fitting of individualized in-ear-EEG devices (Figure 4-1D). Imprints from both ears were taken by trained audiologists (Akademie für Hörakustik, Lübeck, Germany). Earmolds were manufactured by *Oticon* (Oticon A/S, Copenhagen, Denmark). Each earmold was attached with three in-ear-EEG electrodes (see section 2.2).

4.1.3.2 Stimuli & task

A set of 168, one-second long sounds was used in this study (Santoro et al., 2014). A random sound sequence was generated for every subject individually. To reduce predictability, a jittered inter-stimulus interval (ISI) of 0.3 to 3.1 seconds was inserted (Figure 4-1B). Ten percent of the sounds were presented twice in a row, while participants were instructed to press a button as soon as such a repetition occurred (1-back task). Due to technical issues, button presses weren't recorded. However, the task was comparably easy and only assured that subjects were listening to the stimuli. In total, eight blocks of approximately eight-minute length were presented. During presentation, subjects were asked to visually fixate a cross on the screen.

The presentation of sounds was controlled from *MATLAB 2017a* (*The MathWorks, Inc.*, Natick, Massachusetts, United States) and an *RZ6* audio device (*Tucker-Davis technologies*, Alachua, United States). The outputs of the *RZ6* were connected to the direct input of hearing aids (*Oticon A/S*, Smørum, Denmark). The audio outputs of the hearing aids were connected to the earmolds. Triggers at sound onsets were sent to via *labstreaminglayer* (https://github.com/sccn/labstreaminglayer).

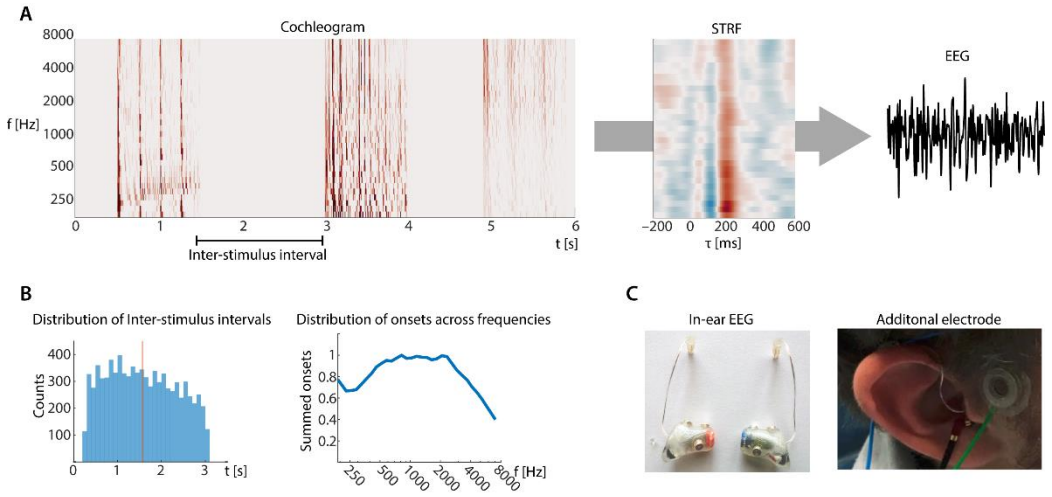

Figure 4-1: Stimulus, prediction, and in-ear EEG configuration. A) Exemplary stimulus representation (Cochleogram) of three successive sounds. Estimated neural responses (STRFs) to the band-wise onsets were used to predict the EEG. B) Left: Distribution of inter-stimulus intervals. Right: Distribution of onsets across frequencies. C) Left: Earmolds with attached in-ear EEG electrodes. Right: Additional electrode attached to the scalp.

4.1.3.3 Data acquisition

Since the goal of this study was to record EEG from a hearing-aid compatible electrode configuration, we reduced the number of EEG electrodes to a minimum. In addition to the 2×3 in-ear EEG electrodes, we attached an electrode in front of each ear at the height of the temple, respectively. The additional electrodes were used as reference for ipsi-lateral in-ear-EEG electrodes. Hence, three channels were recorded per ear. The orientation of the channels should capture activity in temporal brain regions (Figure 4-1C).

Ground and DRL electrode were placed at the forehead of the subjects. All electrodes were connected to a *Smarting* EEG amplifier (*mBrainTrain*, Belgrade, Serbia) via custom-soldered connectors. The comparably small amplifier was attached to a headband. The conductance of non-in-ear electrodes was enhanced using electrolyte gel. Impedances were kept below 20 kΩ.

EEG data and triggers were recorded in one file with *LabRecorder* (*https://github.com/sccn/labstreaminglayer*). The recording of subject 2 stopped due to technical issues, such that only 7 of 8 blocks could be recorded.

4.1.3.4 Extraction of stimulus features

Extraction of stimulus features was conducted using the *NSL-toolbox* (*http://www.isr.umd.edu/Labs/NSL/Software.htm*), which resulted in a 128-channel *cochleogram* covering a frequency range approximately between 200 and 8000 Hz. The 128 logarithmically

spaced frequency channels were down-sampled to 32, 16, 8 (Hann-window with 50% overlap) or summed up to one channel (broad-band). The halfwave-rectified temporal first derivative of the cochleogram resulted in the envelope onsets within each band (Figure 4-1A). The cochleogram summed across time shows the distribution of envelope onsets across the whole frequency range, with a slightly increased density between 500 and 2000 Hz (Figure 4-1B).

4.1.3.5 Data analysis

STRFs were trained by forward-modelling the stimulus-frequency-dependent neural response via ridge regression (See general methods). Beforehand, we cut both the stimulus and the EEG data into one-minute parts, which resulted in approximately sixty parts per subject (59; 46; 60; 58; 57; 58). STRFs were trained in a part-wise leave-one-out fashion and the EEG signal of the left-out part was predicted at every in-ear-EEG channel.

To evaluate the STRF-based prediction, we first calculated the Pearson-correlation coefficient r (further called *neural tracking*) between the predicted EEG signal and the measured EEG signal (i.e., *true*) as well as a randomly chosen EEG signal from another block (i.e., *surrogate*). *Prediction accuracy* was defined as the percentage of parts where the *true* EEG signal reached higher correlation than the *surrogate*. Given that the in-ear EEG channels capture some neural activity related to the stimulation, the STRFs should explain some variance in the EEG signal. Hence, the *true* EEG signal should yield greater correlation-coefficients than the *surrogate* EEG signal. Based on the number of parts, the binomial chance level was calculated per subject based on the chance level of 50%.

4.1.4 Results

In this study, we evaluated the *neural tracking* and *prediction accuracy* of *spectro-temporal response functions* (STRFs) recorded with in-ear EEG. STRFs are estimates of the stimulus-frequency dependent neural response. In-ear EEG is a set of three electrodes placed inside each ear canal, as it could possibly being attached to a hearing aid. We presented stimuli rich of spectro-temporal modulation, while participants were asked to detect repeated sounds (one-back task). We forward-modelled the STRFs and tested their predictive power based on two measures: *Neural tracking* is a measure of how strong the neural response is represented in the EEG signal. *Prediction accuracy* is measure of how precise our prediction fits the measured EEG data

compared to randomly chosen EEG data (i.e., *surrogate*). Due to the low number of subjects (N = 6), we will provide single subject data and apply single subject statistics.

4.1.4.1 In-ear EEG captures neural tracking

First, we looked at the *neural tracking* to the broad-band envelope onsets within single subjects and in-ear-EEG channels. Enhanced *neural tracking* was found for all but one of the subjects (S5) and at all six in-ear EEG channels (Figure 4-2A, first row). The average *neural tracking* varied between channels and subjects, but the *true neural tracking* was generally greater than the *surrogate neural tracking* (except S5). The underlying correlation coefficients spanned a range approximately between 0.01–0.05.

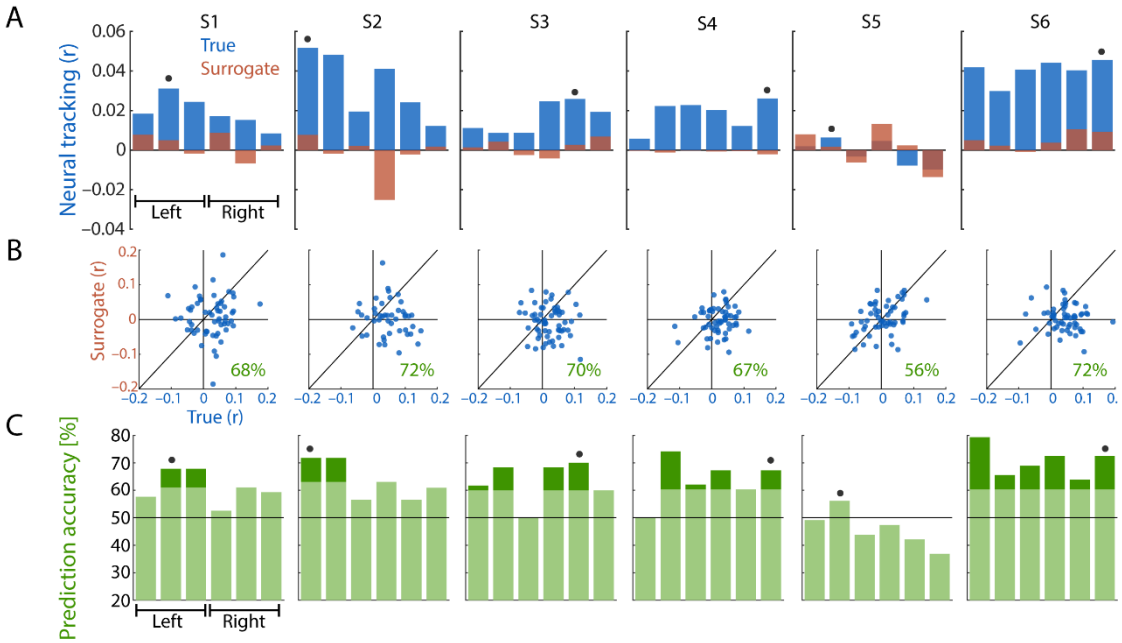

Figure 4-2: ***Neural tracking* and prediction accuracy.** A) *Neural tracking* at six in-ear EEG channels estimated on the measured EEG signal (true) and an EEG signal randomly drawn from another block (surrogate). Black dots indicate the best channel of each subject, which was chosen for further comparisons. B) Scatter plots of estimated *neural tracking*. Percentage indicates prediction accuracy. C) Prediction accuracy as the percentage of trials the true EEG signal yielded higher *neural tracking* than the surrogate EEG signal.

4.1.4.2 Prediction accuracy above chance

In all but one subject (S5) we found at least three out of six in-ear-EEG channels where *prediction accuracy* reached (or exceeded) single-subject chance level (Figure 4-2C; $\alpha = 0.05$, chance = 0.5). This indicates that the modelled TRFs to the onsets of the broad-band envelope were explained stimulus-related variance in the EEG signal.

Figure 4-2B shows the underlying distribution of *true* and *surrogate* correlation-coefficients. Visual inspection suggests that the overall variance of the underlying correlation coefficients is similar across *true* and *surrogate* EEG signals. However, for subjects with higher *prediction accuracy*, the *true* correlation coefficients are more biased towards positivity.

4.1.4.1 Spectro-temporal response functions show frequency dependence

The stimulus features (i.e. onsets) were extracted with different spectral resolution (i.e., 1, 8, 16 and 32 channels). Here we asked if higher spectral resolution leads to an enhanced *prediction accuracy* or if the prediction is mainly relying on the overall broad-band temporal modulation.

Across all spectral resolutions, we found the typical temporal pattern of three succeeding components (P1-N1-P2; Figure 4-3A). Interestingly, the $P2_{STRF}$ was much more prominent compared to other studies within this thesis. In general, the average STRFs look similar across the number of bands. The $N1_{STRF}$ and $P2_{STRF}$ seem to be enhanced towards lower frequencies, indicating a frequency dependent neural response. Towards higher spectral resolutions, The STRFs suggest a more fine-grained, frequency-specific pattern.

4.1.4.2 Higher spectral resolution does not enhance prediction accuracy

Next, we were asking if higher spectral resolution leads to estimations of enhanced *neural tracking* as well as to an enhanced *prediction accuracy*. For all subjects, the estimated *neural tracking* decreased towards higher spectral resolution (Figure 4-3B), which indicates that a spectrally-resolved stimulus representation does not lead to more precise prediction of the EEG signal. Only two subjects (S2 and S6) show and increasing prediction accuracy towards higher spectral resolution.

Unsurprisingly, we observed a strong relationship between the *neural tracking* and *prediction accuracy*, suggesting that *prediction accuracy* is generally depending on the strength of *neural tracking* captured by an (in-ear) EEG electrode. As indicted before, no general trend of enhanced (or decreased) estimations of *neural tracking* or *prediction accuracy* can be found towards higher spectral resolution of the stimulus representation.

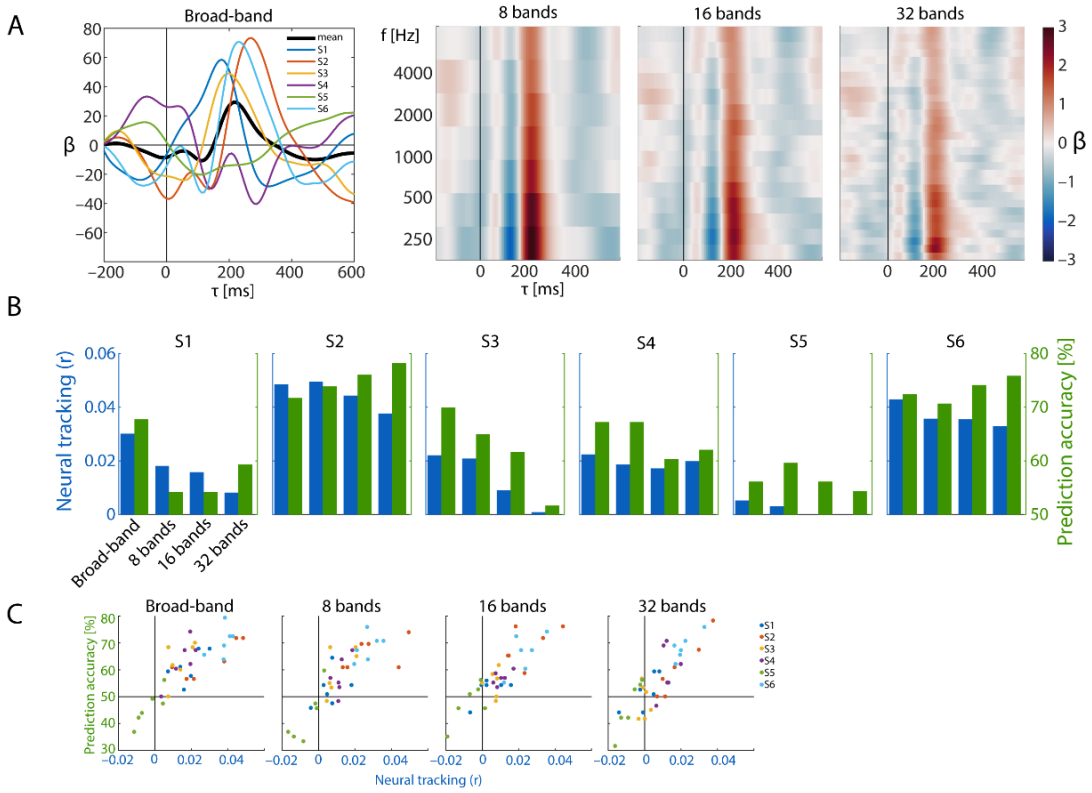

Figure 4-3: Spectro-temporal response functions and prediction accuracy. A) TRF and STRFs to stimulus representations of different spectral resolutions. B) *Neural tracking* and prediction accuracy of single subjects for different spectral resolutions. C) Scatterplots of *neural tracking* versus prediction accuracy at all channels of all subjects.

4.1.5 Discussion

We investigated the stimulus-frequency dependent *neural tracking* captured by single-channel in-ear EEG. After participants listened to a sequence of sounds rich of spectro-temporal modulation, we trained *spectro-temporal response function* (STRFs) on the recorded in-ear EEG signals. We estimated the strength of *neural tracking* and we tested the *prediction accuracy* of the STRFs against random EEG signals unrelated to the stimulus. Here we show that stimulus-frequency dependent neural responses can be extracted from single in-ear-EEG channels. However, increased frequency resolution did not increase the estimated *neural tracking*.

4.1.5.1 The magnitude of neural tracking is comparable to scalp EEG

Neural tracking was found in a range between correlation-coefficients of 0 and 0.08, which is comparable to previous studies (O'Sullivan et al., 2014; Mirkovic et al., 2015; Fiedler et al., 2019; see section 3.1). Even if the explained variance is weak in absolute terms, the observed bias towards more positive correlation coefficients led to *prediction accuracy* above chance in almost all subjects. We thus conclude that the configuration of in-ear EEG electrodes together with an ipsi-

lateral electrode superior frontal to the ear is feasible to capture neural responses from auditory cortical areas.

4.1.5.2 Increased spectral resolution does not lead to a more precise prediction

Contradictory to our hypothesis, a higher spectral resolution of the stimulus representation did not lead to an enhanced estimation of *neural tracking* or *prediction accuracy*. This indicates that the voltage fluctuations recorded from in-ear EEG mainly capture the response to overall, broad-band sound intensity fluctuations. However, the STRFs showed similar components as the broad-band TRFs. A slightly stronger response to lower frequencies was observed, which indicates that some frequency specificity was captured by the in-ear EEG electrodes. Apparently, this frequency specificity did not outweigh the increased number of parameters (i.e., regressors) the model must be fitted to, such that it did not result in more precise predictions. Consequently, the spectral resolution of the stimulus should be chosen only as high as necessary, such that the number of regressors can be kept to a minimum.

Regarding neurally steered hearing aids, it should be further investigated if the frequency-dependent neural responses is modulated by a sensorineural hearing loss. Here, we only investigated the neural response to attended stimuli without any distraction. In particular, it should be investigated how signatures of selective attention are affected by the frequency-dependent amplification of a hearing aid.

4.1.5.3 Conclusion

Our results indicate that the estimation of the frequency-dependent *neural tracking* of auditory inputs could be implemented in hearing aids, but the spectral resolution should be adapted to the current application. If only a general estimate of the neural response is needed, a broad-band representation of the stimulus is adequate. However, if a spectrally-resolved measure is needed, in-ear EEG might be feasible as well, but the spectral resolution should be kept to a minimum.

4.2 Study 5: Single-channel in-ear EEG detects the focus of auditory attention to concurrent tone streams and mixed speech[3]

4.2.1 Abstract

Conventional, multi-channel scalp electroencephalography (EEG) allows the identification of the attended talker in concurrent-listening ('cocktail party') scenarios. This implies that EEG might provide valuable information to complement hearing aids with some form of EEG and to install a level of neuro-feedback. To investigate whether a listener's attentional focus can be detected from single-channel hearing-aid-compatible EEG configurations, we recorded EEG from three electrodes inside the ear canal ('in-ear EEG') and additionally from 64 electrodes on the scalp. In two different, concurrent listening tasks, participants (n = 7) were fitted with individualized in-ear EEG pieces. They were either asked to attend to one of two dichotically-presented, concurrent tone streams or to one of two diotically-presented, concurrent audiobooks. A *forward encoding model* was trained to predict the EEG response at single EEG channels. Each individual participants' attentional focus could be detected from single-channel EEG response recorded from short-distance configurations consisting only of a single in-ear EEG electrode and an adjacent scalp EEG electrode. The differences in neural responses to attended and ignored stimuli were consistent in morphology (i.e. polarity and latency of components) across subjects. In sum, our findings show that the EEG response from a single-channel, hearing-aid-compatible configuration provides valuable information to identify a listener's focus of attention.

4.2.2 Introduction

In multi-talker situations, hearing-aid users find it difficult to comprehend the attended conversational partner against background noise (i.e. cocktail party problem, Cherry 1953). Part of this problem might be caused by the fact that the hearing aid is lacking the explicit information which sound source the listener wants to listen to. The investigation of the *neural tracking* of speech (for a method's review, see Wöstmann et al., 2016) using Electroencephalography (EEG)

[3] *This section is adopted from a published article (Fiedler et al. 2017) with contributions to the study design, analysis and writing from Malte Wöstmann, Carina Graversen, Thomas Lunner & Jonas Obleser.*

and identification of the attended talker in multi-talker scenarios from multichannel scalp-EEG (Mirkovic et al., 2015, O'Sullivan et al., 2014) has demonstrated that EEG could feasibly inform future hearing aid algorithms about a listener's focus of attention. Information about the focus of attention would allow hearing aids for example to adapt noise suppression algorithms or to align directional microphones to the attended sound source (Mirkovic et al., 2016; Van Eyndhoven et al., 2016).

The implementation of EEG into comparably small hearing aids allows the attachment of only few electrodes at restricted positions inside the ear canal (Bleichner et al., 2015; Mikkelsen et al., 2015) or around the ear (Debener et al., 2015, Mirkovic et al., 2016). Since EEG responses quantify the potential difference between a signal electrode and a reference potential, at least two electrodes are required to measure the EEG. The position and distance as well as the orientation of the two electrodes mainly determines, if relevant and irrelevant electrophysiological and external sources will be captured, respectively. Due to the limited number of channels in such a hearing-aid-compatible configuration, established offline methods of EEG signal enhancement such as *independent component analysis* relying on covariance of multiple, whole scalp covering electrodes (Makeig et al., 2004) are not applicable.

An established method to extract auditory evoked potentials (AEP) is based on multiple time-locked presentations of identical stimuli and the subsequent averaging of the measured EEG time-domain signal (Rockstroh et al., 1982). Using this method, it has been shown that the AEP can be extracted from the potential difference between in-ear EEG electrodes and adjacent scalp-EEG electrodes (Bleichner et al., 2015, Mikkelsen et al., 2015, Fiedler et al., 2016). For the presentation of continuous, non-repetitive speech, averaging across multiple trials is not applicable (for review see Wöstmann et al., 2016). Thus, a method to estimate a response evoked by continuous speech is needed. Importantly, the quasi-rhythmic fluctuations of the speech signal's broad-band temporal envelope have recently been reconstructed successfully from Magnetoencephalography (MEG) (Ding and Simon, 2012) and EEG (Mirkovic et al., 2015, O'Sullivan et al., 2014) using linear models. Despite some remaining ambiguities as to the signal features that do get encoded in the neuro-cortical signal (see e.g. Ding and Simon, 2014), a main finding here is that the attended-speaker signal attains a dominant representation in the measured neural signal.

In sum, recent scalp-EEG research has established the feasibility to infer on a listener's attentional focus from EEG very generally. In this present study, however, the overriding goal is to examine single-channel in-ear EEG configurations that possibly could be part of a hearing aid. To this end, we focus our analyses on single-channel electrode configurations consisting of an in-ear EEG and a scalp EEG electrode close to the ear only, to allow future smooth integration with extant hearing-aid systems (Lunner and Gustafsson, 2016). We employ estimation of a forward (i.e. encoding) model since we focused on the encoding of onsets in the broad-band temporal envelope and the prediction of the to-be-expected EEG-signal at single EEG channels. Furthermore, we avoided any methods of artefact rejection such as independent component analysis or trial rejection. This approach allows us to presume that the same results could have been achieved by solitarily recording the respective channel by attaching only two electrodes.

The resulting data from two challenging, cocktail-party-like listening paradigms demonstrate that, on the single participant level, we are able to accurately infer a listener's attentional focus from a single-channel EEG setup consisting of electrodes in and around the ear.

4.2.3 Methods

4.2.3.1 Participants

Eight subjects were enrolled in the study (aged 23, 25, 28, 29, 39, 41, 43 and 49; 4 males). Each participant was provided with individually fitted ear molds. Each ear mold was equipped with three in-ear EEG electrodes (Fiedler et al., 2016; see section 2.2). Five of the subjects were native Danish speakers, while two were French and one was a German native speaker. All reported normal hearing and no histories of neurological disorders. Participants gave informed consent. Procedures were in accordance with the *Declaration of Helsinki* and approved by the local ethics committee of the *University of Leipzig Medical faculty*. All subjects participated in the *oddball task*, while only the five native Danish speakers participated in the *audiobooks task* (aged 29, 39, 41, 43 and 49, 3 males). For both tasks, the recording from one of the Danish subjects had to be discarded due to invalid in-ear EEG data, as the device did not remain in place during recordings. Note that the comparably low number of subjects is caused by the fact that in-ear EEG devices are in a prototype stadium and can't be manufactured in high quantities. However, all results presented are based on rigorous levels of statistical significance in the single subject.

4.2.3.2 Stimuli and tasks

We implemented two experimental paradigms to investigate whether neural responses for two concurrent auditory streams can be extracted from in-ear EEG and whether such responses can predict which out of two streams is being attended.

First, we implemented a non-speech, two-stream, dichotic tone paradigm, in close analogy to Lakatos et al. (2013), hereafter called *oddball task*. Two dichotically presented (i.e. left versus right ear) concurrent streams of 100 ms tones (with a sawtooth carrier waveform) were presented for 1 min. On each trial, the two streams differed in tone repetition rate (1.4 versus 1.8 Hz) and pitch (410 versus 610 Hz). 10–15% of the tones occurred as oddballs (1/4 tone pitch deviation) in both streams. Participants were asked to either attend to the stream presented on the left or right ear and to press a button with their right hand as soon as they heard an oddball in the attended stream. In total, 40 trials of 1 min length were presented (Figure 4-4). All stimulus manipulations, repetition rate (1.4 versus 1.8 Hz), pitch (410 versus 610 Hz), and attention (left versus right) were counterbalanced across trials.

The second paradigm was a two-stream, continuous-speech paradigm, hereafter called *audiobooks task*. Emulating typical challenging listening scenarios, we presented a mixture of two concurrent audiobooks to both ears (i.e. diotic presentation without any spatial cues; Figure 4-4A). The stimuli were two different Danish works of fiction spoken by a female (*Marryatt, Children of the forest*) and a male speaker (*Poe, A Descent into the Maelström*), with matched long-term root-mean-squared (rms) sound intensity. Each exemplar of one-minute mixtures was presented twice in succession. Counterbalanced across trials, subjects were asked to either attend to the male voice first and second to the female voice or vice versa. In total, 60 trials of such one-minute mixtures were presented.

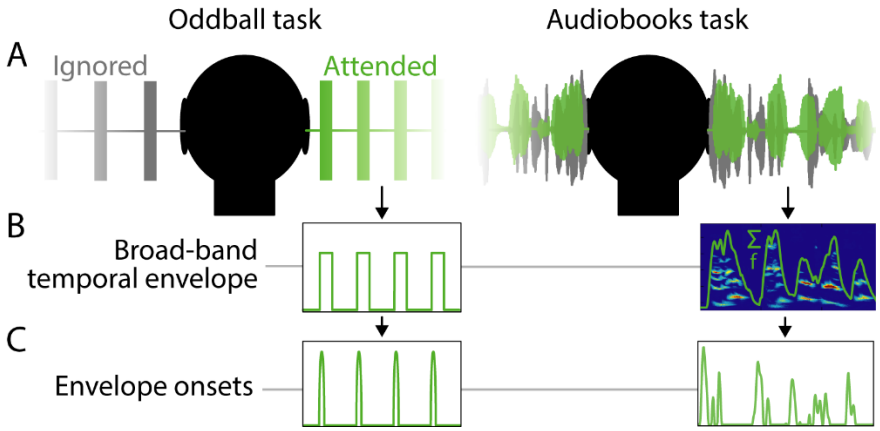

Figure 4-4: Design and Envelope onset extraction. A) Exemplary stimulus waveforms show the spatial separation of target (green) and distractor (grey) stimuli in both tasks. In the oddball task, two streams of 100-ms tones differing in repetition rate and pitch were presented. Subjects were asked to attend to the left or the right stream and press a button as soon as they heard an oddball (pitch deviation) in the attended stream. In the audiobooks task, two Danish audiobooks spoken by a female and male speaker were presented. The identical mixture of both speakers was presented on both ears (diotic). Subjects were asked to attend either the female or the male voice. B) In the oddball task, the broad-band temporal envelope was captured from the stimulus-waveforms directly. In order to capture the broad-band temporal envelope from the audiobooks, an auditory time-frequency representation was summed up across its spectral sub-bands. C) The envelope onsets was obtained by computing the first derivative of the broad-band temporal envelope and subsequently zeroing values smaller than zero (half-wave rectification).

4.2.3.3 EEG-data acquisition and preprocessing

Sixty-four-channel scalp-EEG was recorded alongside in-ear EEG using an *ActiveTwo* amplifier (*Biosemi*, Amsterdam, Netherlands). In-ear EEG electrodes were connected to the auxiliary inputs of the amplifier via pre-amplifiers identical to the ones used for scalp EEG. EEG data were recorded with a sampling rate fs = 2048 Hz. Data were preprocessed using both the fieldtrip toolbox (Oostenveld et al., 2011) for *MATLAB 2017a* (*The MathWorks, Inc.*, Natick, Massachusetts, United States) and custom-written code. The continuous EEG data recorded during the *oddball task* were highpass-filtered at fc = 1 Hz and lowpass-filtered at fc = 15 Hz. The continuous EEG data recorded during the audiobooks task were highpass-filtered at fc = 2 Hz and lowpass-filtered at fc = 8 Hz according to O'Sullivan et al (2014). In order to compensate phase shifts, data were filtered both forward and backward using Hamming window FIR filters with orders N = 3fs/fc. Subsequently, all data were down-sampled to 125 Hz to match the sampling rate of the envelope onsets (see below).

After an initial inspection of the event-related potential (ERP) between in-ear EEG electrodes and *Cz*, we encountered the issue of not all in-ear EEG electrodes keeping proper conductance across the whole experiment. Thus, for each ear canal, only the electrode showing minimal standard deviation across trials in the ERP summed up between 0 and 500 ms relative to tone-

onsets was selected for further analysis. In order to evaluate the potential difference between in-ear EEG electrodes and scalp EEG electrodes, we created two datasets for each participant, one with all scalp channels referenced to the priorly selected left in-ear EEG electrode and the other with all scalp-EEG channels referenced to the selected right in-ear EEG electrode.

4.2.3.4 Extraction of envelope onsets

Several approaches to extraction of the broad-band temporal envelope from a speech signal have been proposed (Biesmans et al., 2016, Thwaites et al., 2016). In case of the *oddball task*, the envelope was extracted by a direct calculation of the absolute values of the analytic signal. In case of broad-band speech signals, the analytic signal is only a rough approximation and it has been shown that an intermediate step of extraction and subsequent summation of frequency sub-band envelopes increases the accuracy of detecting the attended talker (Biesmans et al., 2016; see section 2.3). Thus, for the *audiobooks task*, we extracted the sub-band envelopes using *NSL Toolbox* (Ru, 2001), which resulted in a representation containing the envelopes of 128 frequency bands of uniform width on the logarithmic scale with center frequencies logarithmically spaced between 0.1 and 4 kHz (24 bands per octave). In order to obtain the broad-band temporal envelope, sub-band envelopes were summed up across frequency (Figure 4-4B).

Furthermore, it has been proposed to transform the broad-band temporal envelope in order to extract salient increases of signal power (Hertrich et al., 2012, Hambrook and Tata, 2014). This method is based on the assumption that earliest time points of sensation that could evoke responses are tone or syllable onsets, respectively. It can be calculated by zeroing negative values (halfwave rectification) of the first derivative of the broad-band temporal envelope and results in a pulse-train-like series of peaks. Most salient peaks occur both at tone or syllable onsets (Figure 4-4C). This time-series will be called *envelope onsets*. Recently, we have shown that the cross-correlation of the *envelope onsets* and the EEG-signal results in estimations of the neural response are similar to conventional ERPs obtained by multi-trial averaging (Fiedler et al., 2016).

4.2.3.5 Training EEG response models

A schematic illustration of the approach to identification of the attended speaker is provided in Figure 4-5. In order to evaluate the performance in detection of the attended talker at every

single EEG channel, we first trained a model for each individual participant. The model is a linear mapping of the envelope onsets onto the measured EEG signal (see section 2.5).

We used a well-established form of regularized regression (i.e., ridge regression; Hoerl and Kennard, 1970) to train our model, as ridge regression has been shown to be applicable for predicting neurophysiological signals on the base of stimulus features (forward encoding model) (Santoro et al., 2014; Lalor et al., 2009) as well as reconstructing stimulus features from EEG signals (backward decoding model) (O'Sullivan et al., 2014; Mirkovic et al., 2015). A Matlab-toolbox (mTRF Toolbox) is provided (*https://sourceforge.net/projects/aespa*). As established above, the EEG signal should be independently predicted for every single EEG channel, which is, due to the implementation, inherent of *forward modelling* (Crosse et al., 2016).

In detail, a single-channel encoding model g is the linear mapping of the envelope onsets *s* onto the EEG signal r, which can be expressed as a convolution operation

$$\boldsymbol{r}(\boldsymbol{t}) = \boldsymbol{s} * \boldsymbol{g} = \sum_{\boldsymbol{\tau}} [\boldsymbol{s}(\boldsymbol{t} - \boldsymbol{\tau}) \cdot \boldsymbol{g}(\boldsymbol{\tau})] \qquad \textbf{4-1}$$

where t for $t = 1,2,\ldots,L$ is the sample index of both of the envelope onsets and the EEG signal with length L and τ for $\tau_{min}, \tau_{min}+1,\ldots, \tau_{max}$ is the investigated sample-wise time lag between *s* and *r*. We investigated time lags (between the envelope and the EEG signal) ranging from –100 to 550 ms. In our design, we expect a difference in morphology of the response functions g_{att} *and* g_{ign} (Figure 4-5B), which are models of the responses to the attended and the ignored stimulus envelope onsets s_{att} and s_{ign} (Figure 4-5A). Moreover, we assume that the responses r_{att} and r_{ign} sum up and some noise *n* interferes (Zion Golumbic et al., 2013). Accordingly, we can express the measured EEG signal r_{EEG} (Figure 4-5C):

$$\boldsymbol{r}_{\boldsymbol{EEG}}(\boldsymbol{t}) = \sum_{\boldsymbol{\tau}} [\boldsymbol{s}_{\boldsymbol{att}}(\boldsymbol{t} - \boldsymbol{\tau}) \cdot \boldsymbol{g}_{\boldsymbol{att}}(\boldsymbol{\tau})] + \sum_{\boldsymbol{\tau}} [\boldsymbol{s}_{\boldsymbol{ign}}(\boldsymbol{t} - \boldsymbol{\tau}) \cdot \boldsymbol{g}_{\boldsymbol{ign}}(\boldsymbol{\tau})] + \boldsymbol{n}(\boldsymbol{t}) = \boldsymbol{r}_{\boldsymbol{att}}(\boldsymbol{t}) + \boldsymbol{r}_{\boldsymbol{ign}}(\boldsymbol{t}) + \boldsymbol{n}(\boldsymbol{t}) \qquad \textbf{4-2}$$

Since our goal was to estimate a response model including g_{att} *and* g_{ign} that minimizes the mean-squared error of the subsequent predicted EEG response $\hat{r}_{EEG}$, it can be obtained by the standard matrix operation in regularized regression,

$$\boldsymbol{G} = (\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{S} + \boldsymbol{\lambda}\boldsymbol{m}\boldsymbol{I})^{-1}\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{R} \qquad \textbf{4-3}$$

where S is an L-by-2T-matrix with its columns containing envelope onsets of both the attended s_{att} and ignored s_{ign} stimulus envelope onsets and their time-lagged replications. R is a column

vector of length L containing the measured single channel EEG signal r_{EEG}. The relative regularization parameter λ is first multiplied with m, the mean of the diagonal elements of S^TS (Biesmans et al., 2016). Second, it is multiplied with the identity matrix I and added to the covariance-matrix S^TS. This regularization term λmI prevents overfitting (Crosse et al., 2016), which appeared as high frequent artifacts in the to be estimated response models. The resulting matrix G contains the time-lag-wise response weightings g_{att} and g_{ign} for both the attended and ignored stimulus envelope onsets.

After an initial inspection of the response models, we decided to choose $\lambda = 10^2$. Please note that the greater λ is chosen, the more the term ($S^TS + \lambda mI$) converges to a multiple of the identity matrix, and the influence of covariance vanishes. This would lead to the same results as cross-correlation, which was also shown to be feasible for extracting neural responses (Kong et al., 2014; Fiedler et al., 2016), but doesn't account for potential confounds caused by auto-correlation in the regularized signal. Here we couldn't observe a consistent benefit of regularization, because classification accuracy did not decrease by further increasing λ. However, in order to be consistent with the literature, we applied regression as stated above.

In line with former studies (O'Sullivan et al., 2014; Mirkovic et al., 2015), we decided to apply leave-one-out cross-validation. According to Biesmans et al. (2016) we trained the prediction models by concatenating both the stimuli and EEG signal of all but the to-be-tested trial. Thus, we obtained a prediction model for every single trial.

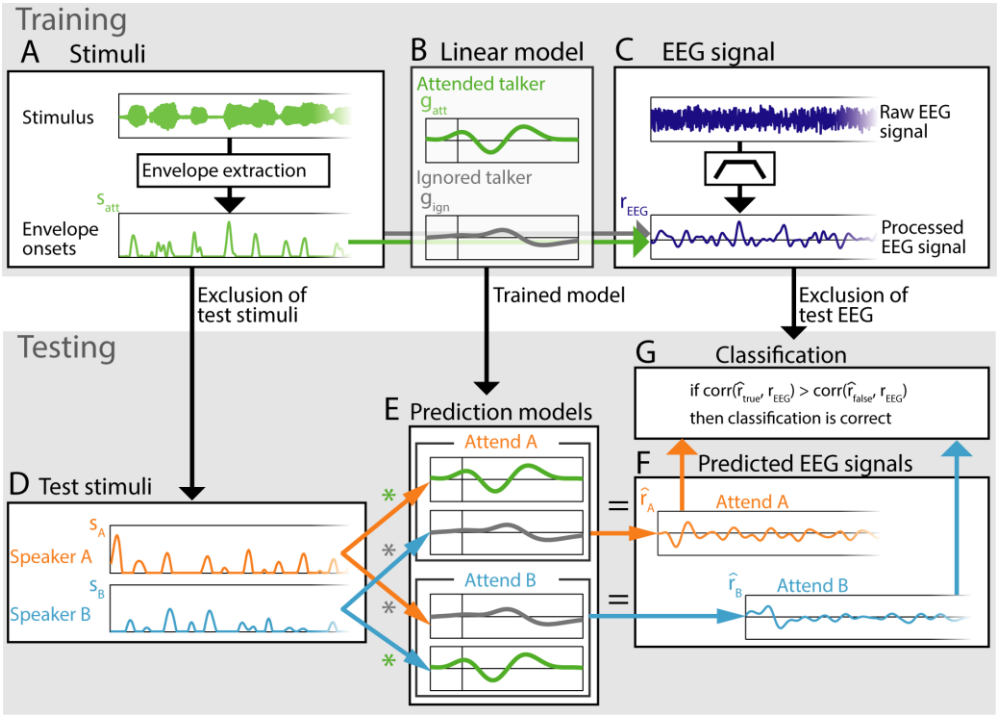

Figure 4-5: Identification of the attended speaker from single-channel EEG exemplary for audiobooks task. Training: After extraction of the onset-envelope (A) and preprocessing of the EEG signal (C), a linear *forward model* (B) is estimated for each trial and each speaker by concatenated stimulus and EEG signal of all other trials. Testing: The convolution of the envelope onsets of speaker A and B (D) with the trained prediction models (E) predicts to be expected EEG signals $\hat{r}_A$ and $\hat{r}_B$ with the labels 'Attend A' and 'Attend B', respectively (F). G) If the predicted EEG signal labeled true (i.e., corresponds to the trial instruction) yields higher Pearson-correlation coefficient with the measured EEG-signal than the predicted EEG signal labeled false (i.e., is contrary to trial instruction), the classification is correct.

4.2.3.6 Testing EEG response models: Identification of the attended stream

In order to classify which of the streams a listener attended to, the former trial-wise trained models g_{att} and g_{ign} (figure Figure 4-5B) were assembled to become two contrary prediction models (Figure 4-5E). According to (1), the sum of the convolution of the envelope onsets s_A and s_B (Figure 4-5D) and each response model (Figure 4-5E) predicts an EEG signal, respectively. For both scenarios with the labels Attend A and Attend B, EEG signals $\hat{r}_A$ and $\hat{r}_B$ (Figure 4-5F) were predicted:

$$\hat{r}_A(t) = \sum_{\tau}[s_A(t-\tau)\cdot g_{att}(\tau)] + \sum_{\tau}[s_B(t-\tau)\cdot g_{ign}(\tau)] \quad \textbf{4-4a}$$

$$\hat{r}_B(t) = \sum_{\tau}[s_A(t-\tau)\cdot g_{ign}(\tau)] + \sum_{\tau}[s_B(t-\tau)\cdot g_{att}(\tau)] \quad \textbf{4-4b}$$

This operation can be expressed by matrix multiplication of the envelope onsets matrix *S* and the response model matrix *G*:

$$\widehat{\boldsymbol{R}} = \boldsymbol{SG} \qquad \textbf{4-5}$$

where $\hat{R}$ is a column vector containing the predicted EEG signal $\hat{r}_A$ or $\hat{r}_{\mathrm{B}}$, respectively.

In order to estimate which of the predicted EEG signals ($\hat{r}_{\mathrm{A}}$ vs $\hat{r}_{\mathrm{B}}$) is most likely representing the trial instruction (attend A vs attend B), we calculated the Pearson-correlation coefficient of the predicted EEG signals ($\hat{r}_{\mathrm{A}}$ and $\hat{r}_{\mathrm{B}}$) and the measured EEG signal r_{EEG}, respectively (L = 7500 samples, Figure 4-5G). The predicted EEG signal that matched the to-be-attended stream (A vs B) was labeled true, the other one was labeled false. The classification was considered correct if the predicted EEG signal labeled true yields greater (i.e., more positive) correlation than the EEG signal labeled false.

4.2.3.7 Goodness of fit

As a measure for the goodness of fit, we will refer to the correlation coefficient obtained from Pearson-correlation of the true prediction and the measured EEG signal. The greater this coefficient, the more of the measured EEG signal's variability would be explained by the response model. Because a convolution is a weighted sum and here the weights are the response models with positive or negative weights at certain time lags, the predicted EEG signals should have the same polarity as the measured EEG signal. Hence, the inspection of the correlation coefficient's magnitude (or square) wouldn't be appropriate. Thus, a greater (i.e. more positive) correlation coefficient indicates the true prediction.

4.2.3.8 Classification accuracy

By classification accuracy we will refer to the percentage of trials in which the predicted EEG signal labeled *true* yields higher correlation with the measured EEG signal than the predicted EEG signal labeled *false*. For statistical analyses, both the correlation coefficients resulting from Pearson-correlation of the *true* and the *false* prediction with the measured EEG signal, respectively, were fisher-z-transformed and called z_{true} and z_{false}. Considering the number of trials and the binary nature of the decision between two alternatives Attend A or Attend B, a single-subject chance level was defined at a level of significance $\alpha = 0.05$ based on a binominal distribution (O'Sullivan et al., 2014, Mirkovic et al., 2016). This resulted in thresholds of 65% for the oddball task (40 trials) and 61.67% for the audiobooks task (60 trials).

4.2.4 Results

The main goal of this study was to identify the attended stimulus stream based on responses at single-channel EEG configurations consisting of one in-ear electrode and one scalp electrode. To this end, we trained forward encoding models in order to predict EEG signals containing the predicted responses to both the attended and the ignored stimulus stream. Two alternative EEG signals representing the scenarios Attend A and Attend B were predicted. The prediction corresponding to the to-be-attended stream was called *true* and the other one *false*. Goodness of fit was quantified by Pearson-correlation coefficient of the *true* predicted and the measured EEG signal. For further statistical analyses, this coefficient was Fisher-z-transformed and called z_{true}, whereas its counterpart z_{false} was equivalently computed by correlation of the false prediction and the measured EEG signal. Our approach to classification relies on the assumption that the *true* prediction better fits the measured EEG signal and thus leads to more positive correlation coefficients than the *false* prediction. Based on that, the percentage of correctly classified trials will be referred to as *classification accuracy*. All plots but the topographic maps are showing data from the exemplary configuration of FT7 referenced to the left in-ear EEG channel.

4.2.4.1 Response functions reveal consistent attention-related differences

Applying ridge regression to obtain *forward models* is known to return response functions comparable to ERPs (Lalor et al., 2009, Fiedler et al., 2016). Beyond that, ridge regression can be applied on data measured during the presentation of continuous stimuli such as speech. The above-mentioned difference between the correlation coefficients z_{true} and z_{false} (see below) has to arise from differences between the response functions of the attended and ignored stimuli.

An inspection of the grand average response functions averaged across subjects in the dichotic *oddball task* (Figure 4-5A) indicated that we extracted components equivalent to a P50-N100-P200 complex. The response functions (Figure 4-5A) suggest an enhanced N100-equivalent component in responses to attended tones, which can be confirmed by the consistent differences of the responses to attended and ignored tones (Figure 4-5C). All subjects show a negative deflection in responses to attended tones at around 160 ms, while all but one of the subjects show a positive deflection in responses to attended tones at around 380 ms. The topographies of the differences at time lag of maximal deflections show a bilateral pattern.

In the *audiobooks task*, a clear P50-N100-P200-equivalent complex could be found in the responses to the attended talker (Figure 4-5B). The responses to the ignored talker show only weak magnitudes and suggest a suppression of the responses to the ignored talker. Compared to the *oddball task*, this is leading to a greater difference between the responses to the attended and the ignored talker (Figure 4-5D). Again, the differences of the single subject's response functions show a consistent pattern with a common negative deflection at a time lag of 130 ms and a later positive deflection at around 250 ms (Figure 4-5D). The topographies of the components at 130 ms and 260 ms both have fronto-central patterns, spreading out towards temporal regions.

In both tasks, we have found response functions that show consistent patterns across subjects. In particular the deflections between responses to attended and ignored stimuli are prerequisites for a single channel classification approach (see above). Most interesting, these deflections could even be recorded at scalp EEG electrodes located close to its in-ear EEG reference electrode.

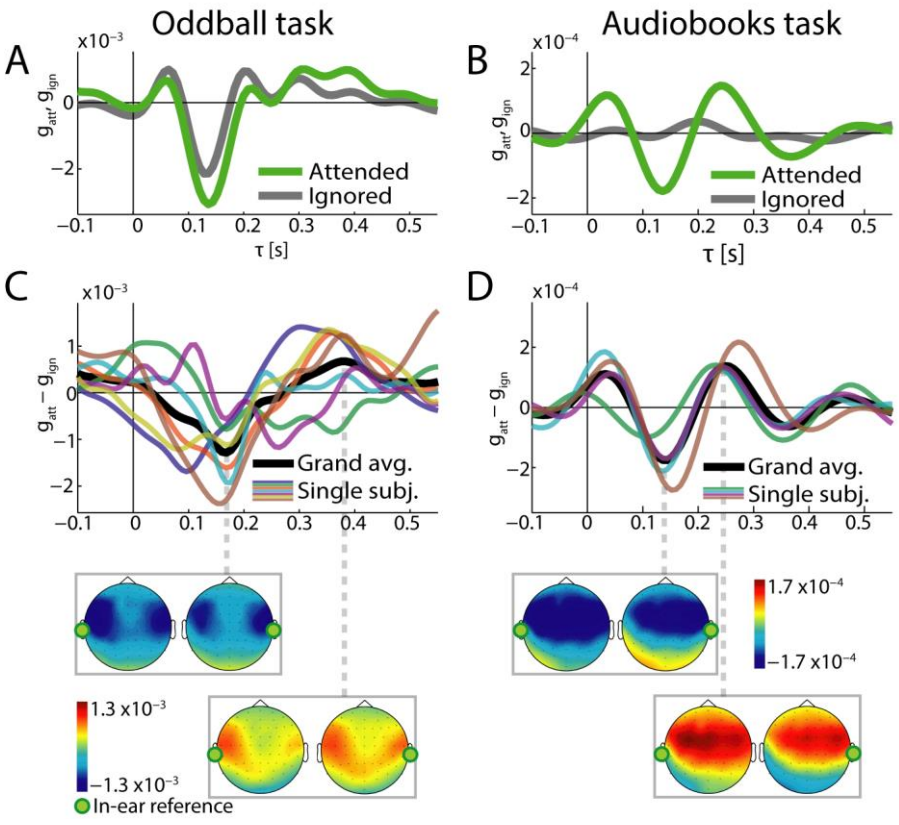

Figure 4-6: Response functions. Response functions shown here were obtained from potential difference between left in-ear EEG and FT7 electrode. A) Grand average response functions to both attended and ignored tones in the oddball task. B) Grand average response functions to both attended and ignored talker in the audiobooks task. C) and D) show single subject data of difference between response functions in the oddball task and in the audiobooks task, respectively. Topographies show grand average weightings at time lags of maximal difference between the response functions (i.e., attended–ignored).

4.2.4.2 Goodness of fit as a basis for identifying the attended stream

Goodness of fit was defined as correlation coefficient resulting from the Pearson-correlation of the measured EEG signal and the predicted EEG signal that consists of the responses to the to-be-attended and to-be-ignored stream (i.e., true prediction).

Generally, the average goodness of fit with values in a range of 0.02–0.15 (oddballs: mean = 0.12, range 0.08–0.15; audiobooks: mean = 0.04, range: 0.02–0.06) seems weak. In order to statistically evaluate if the correlations of the predicted and the measured EEG signals provide valuable information for classification, we investigated the distribution of the Fisher-z-transformed Pearson-correlation coefficients z_{true} and z_{false}. Figure 4-7A & B show the distribution of the correlation coefficients in both tasks, where every single dot represents a single trial performed by a (color-coded) single subject. The correlation of the *true* prediction and the measured EEG signal (z_{true}) tends to be greater than its counterpart z_{false} in the majority of the trials (Figure 4-7A & B). The difference $z_{true} - z_{false}$ was found to be significantly above zero for each subject (one-sample t-test, oddballs: six subjects $p < 0.001$, one subject $p < 0.01$, dof = 39, Figure 4-7C; audiobooks: two subjects $p < 0.001$, one subject $p < 0.01$, one subject $p < 0.05$, dof = 59, Figure 4-7D), suggesting it to be a valuable basis for deciding which of the streams is attended.

In order to evaluate which electrode configuration provides best inference on identification of the attended talker, we inspected the grand average topographies (Figure 4-7C & D) of the single subject t-values obtained from the distribution of the difference between z_{true} and z_{false} (see above). Strongest effects were found at in-ear EEG configurations incorporating fronto-central scalp-EEG channels. Interestingly, in both tasks highest t-values were observed for configurations consisting of scalp-EEG electrodes (i.e. FT7, FT8, T7, T8) close to the ear that the reference in-ear EEG electrode was placed in.

Generally, the analysis of goodness of fit gave insight how a set of two electrodes consisting of one electrode in the ear canal and another at the scalp close to the ear should be oriented in order to explain attention-related variance in the EEG signal caused by auditory stimulation.

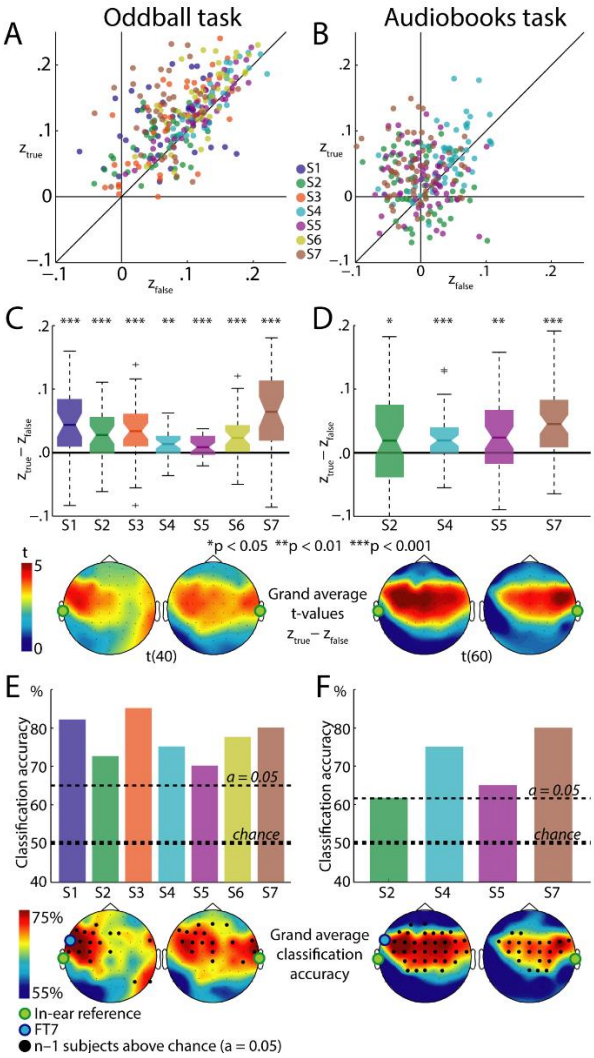

Figure 4-7: Goodness of fit and classification accuracy. Single subject data shown here were obtained from potential difference between left in-ear EEG and FT7 electrode. Topographies show grand average data. A&B) Each dot represents the relation of both Pearson-correlations z_{true} and z_{false} in single trials of the oddball task. C&D) Distributions of the difference $z_{true} - z_{false}$ for single subjects, which were tested against zero (t-test). Topographies show grand average t-values. E&F) Classification accuracy based on the difference $z_{true} - z_{false}$. Horizontal lines indicate significance above chance based in a binominal distribution. Topographic maps show grand average classification accuracy. Highlighted channels are indicating channels where at least n-1 subjects yield classification accuracies significantly above chance.

4.2.4.3 The attended stream can be identified from single-channel configurations

Classification accuracy was defined as the percentage of trials the predicted EEG signal labeled *true* yields a more positive Pearson-correlation coefficient with the measured EEG signal than the predicted EEG signal labeled *false*. For statistical analyses, Pearson-correlation coefficients were Fisher-z-transformed and called z_{true} and z_{false}.

The classification accuracy at FT7 referenced to the left in-ear EEG electrode is shown in Figure 4-7E & F. Classification accuracy was found to be significantly above chance ($p < 0.05$) for all subjects and both the *oddball task* (mean: 77%, range 69–85%, Figure 4-7E) and the *audiobooks task* (mean: 70%, range 62–80%, Figure 4-7F) at this exemplary electrode configuration. Regarding the application in hearing aids, a purely in-ear EEG configuration consisting of two electrodes within the same ear canal is most desirable. We investigated those configurations as

well and provided the results in the supplements (Fiedler et al., 2017; figure S2). Note that these alternative configurations did not yield classification accuracy consistently above chance.

Grand average topographies of classification accuracy (Figure 4-7E & F) show patterns similar to the t-value topographies above (Figure 4-7C & D). Highlighted channels in Figure 4-7E & F indicate that classification accuracy was found to be above chance ($p < 0.05$) for at least all but one of the subjects. Interestingly, channels close to the ear the reference in-ear EEG electrode was placed in showed classification results above chance consistently across subjects.

Due to the low number of subjects, drawing a general conclusion on the most appropriate electrode configuration is not possible. However, for the present data we can state that we have found a configuration, showing classification results above chance for every subject consisting of only two electrodes, FT7 referenced to left in-ear EEG electrode. Single-subject topographical maps provided in the supplements (Fiedler et al., 2017; Figure S1 A) confirm that various short-distance electrode configurations yield classification accuracy above chance. Based on the single-channel data of subjects who participated in both tasks, we found a strong dependency of classification accuracy between tasks (figure S1 B), which emphasizes the robustness of our findings despite our relatively low number of participants.

4.2.5 Discussion

It is a frequently stated long-term goal to fuse EEG recordings with hearing aid technology in order to attune the hearing aid to an attended sound source. Here, we investigated whether the attended sound stream out of two concurring streams can be identified from single-channel EEG recordings. Single channels were electrode configurations consisting of one reference in-ear EEG and one scalp EEG electrode. We focused our analyses on a configuration consisting of a left in-ear EEG electrode and scalp-EEG electrode FT7.

Participants performed two tasks. In both tasks, concurrent sound streams (i.e. tones and speech) were presented. We hypothesized single channel in-ear EEG data to provide valuable information to identify the attended stream.

4.2.5.1 Response functions consistently reveal listeners' focus of attention

In contrast to *backward models*, the estimation of *forward models* allows the comparison of the obtained response functions with conventional ERPs (Lalor et al., 2009). An attention-related difference between response functions is a prerequisite for identification of the attended speaker (see Methods).

In both tasks, we have found an enhanced N100-equivalent component in the responses to attended stimuli compared with ignored stimuli for each subject (Figure 4-6A & B). This is in line with auditory evoked potential (AEP) studies, showing that the N100 component is enhanced if the stimulus is attended (e.g., Näätänen et al., 1981).

Notably, attention-related differences in the response functions could be found even in short-distance configurations consisting of a reference in-ear EEG electrode and a scalp-EEG electrode close to the ear, as exemplarily shown for FT7 referenced to left in-ear EEG electrode. Regarding hearing aid applications, these findings encourage the attachment of only a few electrodes in the periphery of the ear (Mirkovic et al., 2016).

The consistent morphology of the difference between responses to attended and to ignored stimuli (Figure 4-6C & D) further suggests the training of a model based on the data of all but one subject and test it on the latter (i.e., generic model). Even if not as accurate, O'Sullivan et al., (2014) showed that a generic model still allows predicting the attentional focus. With respect to its application in hearing aids, a generic model could provide a default set of parameter values before a listener-specific model is adapted over time (Mirkovic et al., 2015). In the current study, the training of a robust generic model was hindered by the low number of subjects and should be further investigated.

The dichotic oddball paradigm employed here also is appropriate when investigating neural responses to discrete and spatially separated stimuli. However, such a paradigm is removed from real-world listening scenarios, since two or more sound sources in natural environments are rarely separated in a dichotic fashion and are rarely as stationary regarding their rhythm and spectral content.

In contrast, the audiobooks paradigm with two diotically presented talkers represents a challenging listening situation and is more akin to realistic scenarios (also with respect to a

listener's goal, that is, following a sound source and comprehending what is being conveyed (Obleser, 2014). Since no spatial information is contained in the audio signal, a 'worst case' scenario was presented. Sound source separation can only be achieved based on spectral-temporal cues of the two talkers. Since each participant attended to either the male or to the female voice in the same number of trials, the revealed differences of the response function cannot be explained by spatially separated stimuli nor from talker specific features.

In most of the cited studies on detection of auditory attention from EEG data, the speech envelope was used as stimulus representation (O'Sullivan et al., 2014, Mirkovic et al., 2015, Biesmans et al., 2016). In contrast we used envelope onsets, that is, the halfwave-rectified first derivative of the envelope. Using instead the envelope led to similar detection accuracies (Fiedler et al., 2017; figure S4A), but responses were shifted by approximately 50 ms such that the P50 equivalent component appeared before time lag of zero (figure S4 B). This is due to every onset being followed by a peak in the envelope after approximately 50 ms (Fiedler et al., 2017; figure S4 C). For the *oddball task*, the correct latencies of the components (i.e. P50, N100, P200) are known from previously calculated ERPs (Fiedler et al., 2016). Since the latencies of the envelope onsets responses in the audiobooks task fit the latencies of the ERP onset responses in the oddball task better than the envelope responses do, we conclude that envelope onsets lead to more precise estimations.

A comparison of the response functions reveals similar latencies of components between tasks, but the relative suppression of the response to the ignored stream is stronger in the audiobooks task. Two diotically presented talkers are more likely masking each other than dichotically presented tones of 100 ms length (and up to 614 ms pauses between tones). The suppression of the responses to the ignored talker might indicate higher demand for suppression of the ignored stream and thus a higher task difficulty.

Of course, the low number of individually in-ear-fitted subjects tested here (n = 7 & n = 4) allows only for limited conclusions. However, the markedly consistent morphologies of the response functions and the individually significant detection success suggest that differential responses to attended and ignored auditory stimuli, even continuous speech, can be recorded from short-distance electrode configurations. These configurations here consisted only of one electrode in the ear canal and another close to the same ear, as exemplarily shown in Figure 4-7E

& F for a left in-ear EEG electrode referenced to scalp-EEG electrode FT7. Please note that the shortest distance we could achieve was determined by the electrode positions of the scalp EEG. The exemplary electrode FT7 is placed at a distance of approximately 8 cm to the entrance of the ear canal (tragus) at an angle of 40° relative to the tragus-Cz-line. With the development of adhesive electrodes to be attached around the ear it was shown that responses could be recorded at even closer positions (Bleichner et al., 2016).

4.2.5.2 Goodness of fit provides basis for identification of the attended stream

Former studies about approaches to identification of the attended talker mainly used *backward decoding models* (O'Sullivan et al., 2014, Mirkovic et al., 2015, 2016, Biesmans et al., 2016). *Backward models* are trained on multi-channel EEG data and used to reconstruct a single speech envelope. In contrast, we used *forward models* to predict the EEG signal in response to the stimulus, which allowed us to quantify the goodness of fit at every single EEG channel (see section 2.5.1).

The goodness of fit was quantified by Pearson-correlation coefficient for the predicted versus the measured EEG signal. In the previous *backward model* studies cited above, correlation coefficients obtained from Pearson-correlation of the reconstructed and the original speech envelope between 0.02 and 0.10 were reported. Here, we obtained correlation coefficients of similar magnitude, but they were here obtained solely on the basis of a potential difference recorded at a single EEG-channel consisting of left in-ear EEG and scalp-EEG electrode FT7. Crucially, the topographies of single-trial-derived t-values (Figure 4-7C & D) show that meaningful differences can be found satisfyingly at single electrodes close to the referenced in-ear EEG electrode.

We thus conclude that short-distance electrode configurations like the exemplary configuration consisting of the left in-ear EEG reference and FT7 electrode capture information about the listener's attentional focus and thus provide a basis for the identification of the attended sound source. To achieve this, we based our analyses on certain assumptions. First, we assumed that strongest responses can be found at stimulus onsets and thus extracted respective representations (see Methods). Especially for speech, features known to evoke responses are manifold and rarely mutually exclusive, since all are, to some extent, nested or derived from the broad-band temporal envelope (Ding and Simon, 2014). Second, we applied ridge regression in

order to train a model under the assumption of linearity and with the goal to reduce the mean squared error of the prediction. The extraction of features from speech is wedded to the selection of an appropriate model and both affect the contrast between responses to attended and ignored speech.

Comparing several methods of extracting features of speech and going beyond the simple assumption of linearity as well as incorporating several loss-functions might further boost the contrast between the two predicted EEG signals and thus further refine the information about the attentional focus.

4.2.5.3 The attended stream can be identified from single-channel configurations

The major goal of this study was to identify the attended sound stream based on single-channel hearing aid-compatible EEG channel configurations. Considering that, classification accuracy is the most important measure to evaluate the performance of our approach of single channel classification.

As stated above, former studies have used *backward models* to bring in the advantage of having multiple EEG signals to reconstruct one single speech envelope. In order to reduce the number of channels, Mirkovic et al. (2015) already applied an approach of recursive channel elimination. Starting from a grid of 96 channels, it was shown that a stepwise exclusion of worst performing channels doesn't affect classification accuracy up until approximately 25 channels were left. The best performing electrodes were concentrated at temporal positions close to the ear. However, the average of all electrodes served as reference potential which hinders a conclusion for single channel configurations consisting of only two electrodes. In a recent study (Mirkovic et al., 2016), it was shown that based on the data of a grid of ten electrodes around the ear the attended talker could be identified with a *backward model*. Here, we go even further and show that a montage of only two electrodes, left in-ear EEG electrode and scalp-EEG electrode FT7, is sufficient to identify the attended sound source in two experimental tasks. In Mirkovic et al., (2016), we presume that placing a few electrodes at positions favorable for identifying the attended speaker is more crucial than obtaining more or less redundant EEG signals from multiple channels.

With respect to the long-term goal of controlling a hearing aid in real-time, our results provide valuable insight. First, in a hearing aid, computational resources are limited. We thus decided not

to apply any method of artifact rejection or other methods of signal enhancement other than band-limiting the EEG-signal. Once a model is trained, the algorithm consists of only four convolutional operations and two correlations. Considering the comparably low sampling rate of 125 Hz and one-minute trials of 7500 samples, the computational effort is comparably low.

Nevertheless, a classification accuracy of around 70% after one minute might not yet comply with the requirements of a hearing-aid user. Furthermore, data were recorded in a shielded room which reduced environmental noise as well as subjects were asked to move as less as possible which lead to a minimum of muscle artifacts. Please note that an implementation of such an electrode configuration into a hearing aid would raise further issues not addressed here, such as how to attach an electrode outside the ear canal and dealing with low conductance due to hairy positions and skin resistance. One possible solution might be permanently or daily placed electrodes around the ear (Debener et al., 2015, Mirkovic et al., 2016, Bleichner et al., 2016). Thus, for real-life applications, there are still major challenges ahead. Our findings however do map out a significant step towards the application of single channel in-ear EEG in future hearing aids.

4.2.5.4 Conclusion

The identification of attended sound sources based on neural data has become increasingly important for both, neuro-scientists and hearing aid developers, since it contains the potential to control a hearing prosthesis in a brain–computer interface fashion. One unsolved problem is the embedding of EEG electrodes and utilization of EEG signals in the hearing-aid periphery.

In the current study, we have shown that in-ear EEG can feasibly capture information about the listeners' attentional focus. Thus, with only two electrodes attached, an auditory brain-computer interface could constantly track a listener's attentional focus. This information could be fed back to other hearing aid algorithms in real-time (e.g., controlling for directional microphones and noise suppression) at low computational cost.

4.3 Study 6: In-ear EEG detects the focus of auditory attention under continuously varying listening conditions

4.3.1 Abstract

In-ear EEG was previously shown to detect the listener's focus of attention during a diotically presented, concurrent speech scenario where the talkers were matched in sound intensity. However, real-world listening scenarios are not as constant, for example, the signal-to-noise ratio (SNR) varies over time. For scalp EEG, we showed that under negative SNRs additional parietal components in the neural response to the ignored talker indicate enhanced *neural selectivity*. Here we asked, whether the attentional focus under a continuously varying SNR can be detected with in-ear EEG as well. Furthermore, we asked if the parietal component in the response to the ignored talker can be observed in in-ear EEG as well. By running the identical protocol as in our earlier scalp-EEG study, we varied the SNR between two talkers while participants equipped with fitted in-ear EEG listened to one of them. Here, we replicated our earlier findings by showing that in-ear EEG configurations are feasible to detect a listener's focus of attention. Additionally, the modulation of *neural selectivity* between SNRs indicates a late contribution of the ignored talker under the most adverse SNR. This has implications on the development of neurally steered hearing aids, which might profit from an SNR-dependent training of the underlying neural responses.

4.3.2 Introduction

Within this decade, multiple studies showed that EEG signals are informative of a listener's focus of attention in two or more talker scenarios. Those findings encouraged interdisciplinary researchers to develop a hearing aid that is informed about the listener's focus of attention by EEG signals. This endeavor required two questions to be answered: First, which EEG signatures of auditory attention are omnipresent across the manifold of real-world listening scenarios. Second, which EEG electrode configurations are needed to capture those signatures of auditory attention.

Earliest studies about the *neural tracking* of continuous speech used two talkers matched in long-term sound intensity and some of the studies presented dichotic speech (Ding and Simon, 2012; O'Sullivan et al., 2014; Mirkovic et al., 2015). Those conditions are well-suited to investigate the principal mechanisms of attention-dependent *neural tracking* of speech, but only allow

limited conclusion on real-world listening scenarios. Assuming a hearing loss degrades binaural cues for spatial stream segregation, a diotic listening reflects the worst case in terms of spatial segregation. Furthermore, a varying SNR as found in real-world listening scenarios may evoke signatures of *top-down* attentional control that avoid the capture of *bottom-up* attention (see section 3.1). A neurally steered hearing aid may profit from being trained to the SNR-dependent modulation of the neural responses.

Here, we diotically presented two talkers, while the signal-to-noise ratio varied continuously and unpredictably. We trained *forward models* to predict the neural response at single-channel in-ear EEG configurations. We show that a listener's focus of attention can be detected under varying SNR and that late *neural selectivity* of the ignored talker under negative SNRs seems to be detected at in-ear EEG electrodes as well.

4.3.3 Methods

4.3.3.1 Participants

Six native speakers of Danish (3 females) were invited (age: 42, 30, 40, 50, 41, 27). All reported normal hearing and no histories of neurological disorders. All participants gave informed consent. Each participant was provided with individually fitted ear molds with attached in-ear EEG electrodes. Three of the subjects also participated our previous study (Fiedler et al., 2017).

4.3.3.2 Stimuli & Task

We used an experimental design which was identical to our earlier study (Fiedler et al., 2019; section 3.1). In brief, we presented two Danish audiobooks simultaneously read by a female and a male voice, respectively. The identical mixture of both talkers was presented at each ear (diotic). The signal-to-noise ratio (SNR) between the attended (signal) and the ignored (noise) talker varied stochastically between −6 and +6 dB. Twelve blocks of five-minute length were presented, while participants were asked to listen to the female and male voice the same amount of time. The female and male blocks were presented in alternation. After each block, the participants were asked to answer a multiple-choice question (four possible answers) concerning the content of the to-be-attended audiobook. For all subjects, the amount of correctly answered questions was above the chance level of 25% (67%; 92%; 67%; 67%; 92%; 92%). After the questions, the participants

self-determined the length of the break between blocks by starting the next block by pressing a button.

4.3.3.3 Data acquisition and preprocessing

Sixty-four-channel scalp-EEG was recorded alongside in-ear EEG using an *ActiveTwo* amplifier (*Biosemi*, Amsterdam, Netherlands). In-ear EEG electrodes were connected to the auxiliary inputs of the via pre-amplifiers identical to the ones used for scalp-EEG electrodes. EEG data were recorded with a sampling rate fs = 2048 Hz. Data were preprocessed using both the fieldtrip toolbox (Oostenveld et al 2011) for *MATLAB 2017a* (*The MathWorks, Inc.*, Natick, Massachusetts, United States) and custom-written code. The continuous EEG data were highpass-filtered at fc = 1 Hz and lowpass-filtered at fc = 10 Hz (two-pass Hamming window FIR, filter order: 3fs/fc). In order to compensate phase shifts, data were filtered both forward and backward using Hamming window FIR filters with orders N = 3fs/fc. Subsequently, all data were down-sampled to 125 Hz.

In order to evaluate the potential difference between each in-ear EEG electrode and all scalp-EEG electrodes, we created eight datasets for each participant. In six datasets all EEG channels were referenced to an in-ear EEG electrode and in the other two datasets all EEG channels were referenced to the left and right mastoid, respectively (Fiedler et al., 2017). Henceforth, we compared the neural measures between the reference in-ear EEG electrodes and the conventional mastoid reference electrodes.

4.3.3.4 Detection of the attentional focus: neural selectivity

To investigate if we can replicate two of our recent studies (Fiedler et al., 2017; Fiedler et al., 2019; see section 3.1 & 4.2), in a first step, we split up the EEG and stimulus data into 60 blocks of one minute each. In a leave-one-out fashion, we first predicted the EEG signal of each block based on a response model (*temporal response function*, TRF; Crosse et al., 2016) trained on all other 59 blocks, respectively. The TRF contained the neural response to the attended and the ignored talker, respectively. By interchanging the TRFs to the attended and the ignored talker, we predicted a second EEG signal, which did not represent the task instruction (i.e., attend female or attend male). If the first predicted EEG signal yielded higher correlation with the measured EEG signal than the second predicted EEG signal, the attended talker was successfully detected. The

percentage of trials where the attended talker was correctly identified was called *neural selectivity*, as it reflects selective neural processing of the attended versus the ignored talker.

4.3.4 Results

4.3.4.1 Replication: The listener's focus of attention can be detected at in-ear EEG electrode configurations

We identified the attended talker by prediction of EEG signals based on *temporal response functions* (TRFs). The percentage of correctly classified blocks is a measure of how strongly TRFs are modulated by attention and was called *neural selectivity*.

We found enhanced *neural selectivity* at almost all in-ear EEG electrode configurations consisting either of a left in-ear EEG electrode and FT7 or a right in-ear EEG electrode and FT8 (Figure 4-8A). All but S3 showed above-chance *neural selectivity* in at least two of three in-ear EEG electrodes per ear. S3 showed poor *neural selectivity* not only for in-ear EEG configurations but also for conventional reference electrodes, which might be due to technical issues since subject three correctly answered 67% of the questions (chance: 25%). Hence, we excluded the data of subject three from further group analysis. *Neural selectivity* at in-ear EEG electrode configurations compares well to the conventional reference electrodes (left & right mastoid). For further comparisons, we selected the in-ear EEG channel with second best (i.e., median) *neural selectivity* per ear and per subject.

Across the whole scalp, mainly fronto-central electrodes show enhanced *neural selectivity* (Figure 4-8B) both for in-ear and mastoid reference electrodes. The TRFs show the typical pattern of slectively enhanced N1 and P2 in the response to the attended talker compared to the ignored talker (Figure 4-8C). The TRFs are markedly consistent across subjects. Subject three, who did not show enhanced *neural selectivity*, shows the weakest magnitude in the difference between the TRFs to the attended and the ignored talker.

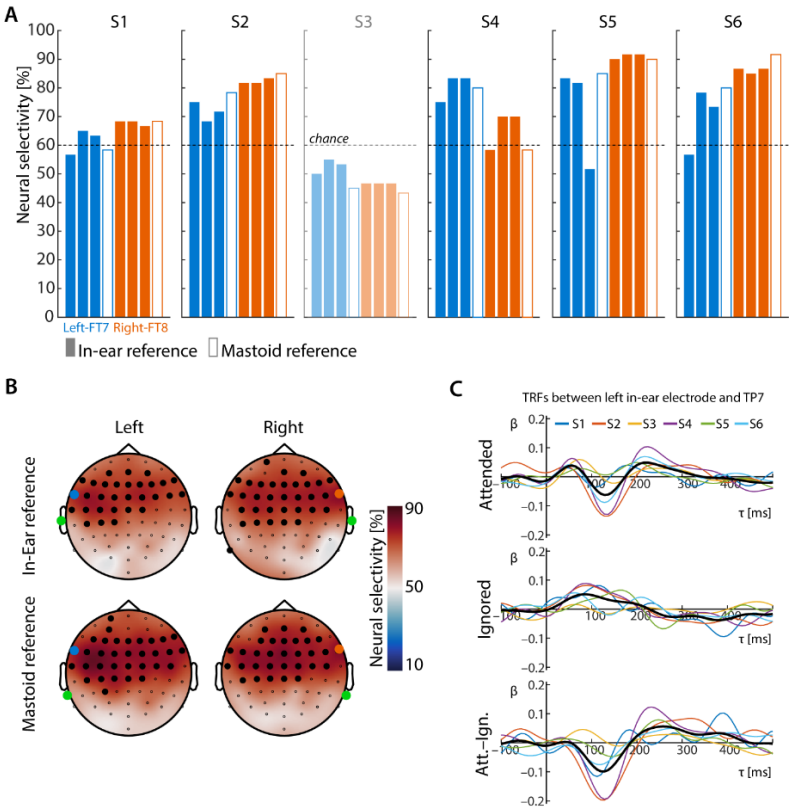

Figure 4-8: *Neural selectivity* and *temporal response functions* (TRFs). A) *Neural selectivity* at all single-channel in-ear EEG and mastoid electrode configurations (left electrodes served as reference for FT7, right electrodes served as reference for FT8; see also B). The in-ear EEG electrode showing median (i.e., second best) *neural selectivity* was selected per ear of each subject for further analysis. Subject three was excluded from further group analysis. B) Group level average of *neural selectivity*. Highlighted channels exceeded the binomial chance level of 60% (Combrisson and Jerbi, 2015). C) TRFs to the attended and the ignored talker at FT7 referenced to the left in-ear EEG electrode.

4.3.4.2 Replication: Late cortical tracking of ignored speech in in-ear EEG

We extracted the TRFs for the three different SNRs to contrast the attention-dependent responses to an acoustically *dominant* and *non-dominant* talker (i.e., –6 vs. +6 dB; see section 3.1; Fiedler et al. 2019). In detail, our goal was to replicate our earlier finding of a louder ignored talker being selectively processed during later time lags of the TRF, which we previously observed at parietal EEG channels.

We found enhanced *neural tracking* and *neural selectivity* comparable to our earlier findings (Figure 4-9, see section 3.1). In tendency, the attended talker is more strongly tracked compared to the ignored (Figure 4-9A). However, under an SNR of –6 dB, the enhanced *neural tracking* of the attended talker was reduced (Figure 4-9A, middle). We observed a late increase of *neural tracking* of the ignored talker in all subjects (Figure 4-9C), which was predominantly found at central and occipital scalp EEG channels, as exemplarily shown for the left in-ear reference. This

enhanced *neural tracking* was accompanied by enhanced *neural selectivity* of the ignored talker, which was found in all subjects mainly at frontal but also occipital channels (Figure 4-9D). Overall *neural selectivity* was found to be robust across the three levels of SNR (Figure 4-9E, left).

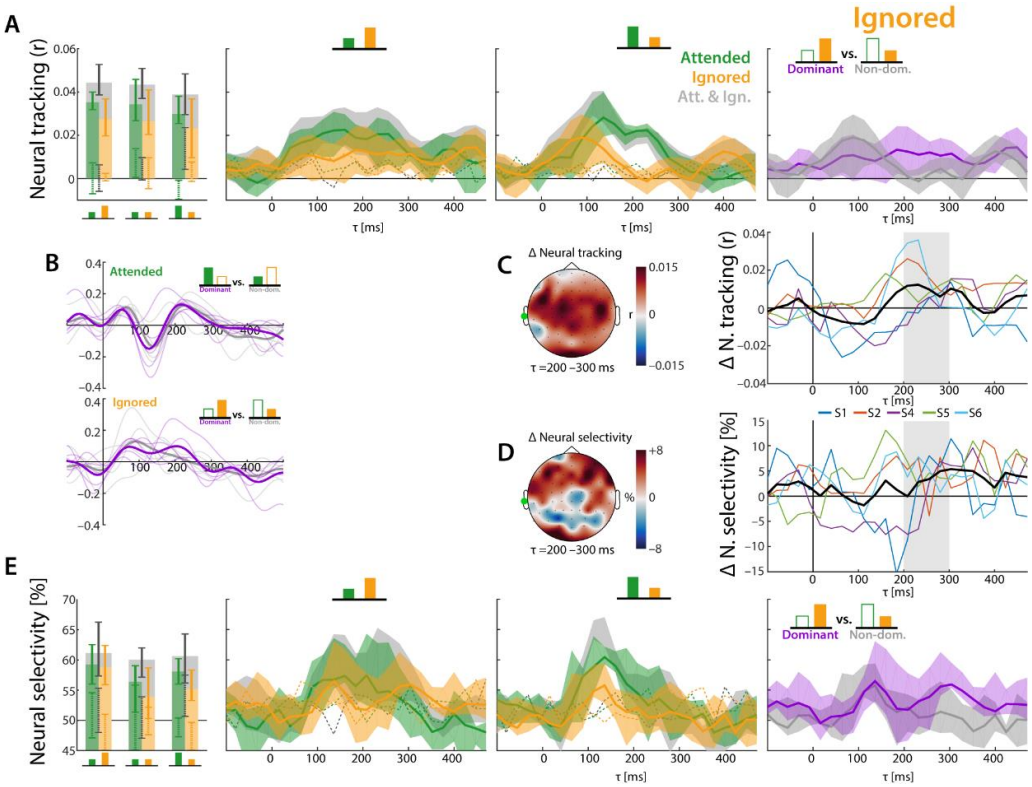

Figure 4-9: *Neural tracking* and *neural selectivity* obtained at scalp EEG electrodes referenced to the left in-ear EEG electrode. Dotted lines represent surrogate data. A) *Neural tracking* of the attended (green), the ignored (orange) and both (grey) talkers. Left: *neural tracking* over all time lags. Middle: time-lagged *neural tracking* during SNR of –6 and +6 dB. Right: *Neural tracking* of the ignored talker during SNR of –6 (purple) and +6 dB (grey). B) Group average TRFs and single subject TRFs (thin lines). C) Left: topographic map of the difference (SNR of –6 vs. +6 dB) of late *neural tracking* of the ignored talker. Right: Difference (SNR of –6 vs. +6 dB) of *neural tracking* of the ignored talker. D) Left: topographic map of the difference (SNR of –6 vs. +6 dB) of late *neural selectivity* of the ignored talker. Right: difference (SNR of –6 vs. +6 dB) of *neural selectivity* of the ignored talker. E) *Neural selectivity* of the attended (green), the ignored (orange) and both (grey) talkers. Left: *neural selectivity* over all time lags. Middle: time-lagged *neural selectivity* during SNR of –6 and +6 dB. Right: *neural selectivity* of the ignored talker during SNR of –6 (purple) and +6 dB (grey).

4.3.5 Discussion

This study was conducted to replicate two of our earlier studies in combination. First study, we showed that single-channel in-ear EEG is informative of a listener's focus of attention (Fiedler et al., 2017; see section 4.1). In our second study, we showed that under negative signal-to-noise ratio (SNR), a late *neural tracking* of the ignored talker contributes to overall selective processing (Fiedler et al., 2019; see section 3.1). Here, we investigated if the attentional focus of a listener can be detected with in-ear EEG under a varying SNR and if enhanced *neural tracking* of the ignored talker under negative SNRs can be found as well.

In sum, we replicated our earlier study showing that a listener's focus of attention can be detected from single channel configurations consisting of in-ear EEG electrodes and short-distant, neighbored scalp EEG electrodes. Again, we highlighted that an electrode configuration pointing towards fronto-central regions is necessary to achieve above-chance *neural selectivity*. Here we additionally showed that this can even be achieved under a varying SNR. Furthermore, our results suggest that there is enhanced *neural tracking* of the ignored talker as well to be found in in-ear EEG configurations.

4.3.5.1 Detection of the attended talker is robust to varying conditions

We showed that even under a varying SNR, the detection of the attentional focus is robust and markedly above chance (Figure 4-8). This has implications on the integration of EEG into neurally steered hearing aids (Lunner and Gustafsson, 2016; Fiedler et al., 2017; Mirkovich et al., 2016). Real-world listening scenarios are rarely constant, since sound sources are varying in intensity and location. Based on our findings, we conclude that within a certain SNR range, a listener's focus of attention can be reliably detected from in-ear EEG, even without training of SNR-dependent neural response models.

We did not include spatial modulation here, which was shown to affect the topography of the neural responses phase-locked to speech as well (Das et al., 2016). The effect of SNR and location on the phase-locked neural response to the talkers should be investigated in a combined way (e.g., section 3.3), since there might a strong interaction to be observed due to a varying spatial release from masking (Schubert and Schultz, 1962; Levitt and Rabiner, 1967).

4.3.5.2 Accounting for SNR-modulated neural responses might improve detection of the attentional focus

As in our earlier study (see section 3.1; Fiedler et al., 2019), we showed that the phase-locked neural response to continuous speech is modulated by the signal-to-noise ratio (SNR). In particular, we found enhanced *neural tracking* and *neural selectivity* of the ignored talker under a negative SNR. That means that the detection of the attentional focus might profit from the training of SNR-dependent response models (i.e., TRFs). Furthermore, the late tracking of the ignored talker might indicate enhanced effort in the selective processing of the auditory scene,

which could be a valuable parameter for the neural steering of hearing aids. This point will be further addressed in a general discussion (see section 5.4).

4.3.5.3 Conclusion

We showed that a listener's focus of attention under varying acoustic conditions can be robustly detected from single-channel in-ear EEG configurations. We also showed that the SNR-dependent modulation of *neural selectivity* revealed a late increase of *neural selectivity* in the response to the ignored talker. In sum, our findings confirmed that in-ear EEG reliably detects a listener's focus of attention and shows that the SNR-dependent modulation might be a valuable resource for a hearing aid to be informed about a listener's attentional state.

5 General Discussion

This thesis comprises six studies which investigated neural signatures of auditory selective attention in the human electroencephalogram. The goals of those studies were two-fold: first, the dynamics of neural signatures of auditory attention were investigated under varying listening conditions, which aimed at revealing neural strategies of *top-down* attentional control to prevent *bottom-up* capture of attention. Second, the recording of neural signatures of auditory selective attention at a reduced set of EEG electrodes (including in-ear EEG) was investigated, which aimed at proofing the feasibility to neurally steer a hearing aid based on those signatures.

5.1 Summary of experimental results

In studies 1–3, we investigated how the neural response to continuous speech is shaped by *bottom-up* and *top-down* attention (see chapter 3). We focused or analysis on neural signatures previously found to be modulated by auditory selective attention: the *neural tracking* of speech and the modulation of *alpha power*. The adversity of the listening condition was manipulated by the variation of the signal-to-noise ratio (SNR) as well as the location of the talkers.

Study 1 showed that the *neural tracking* of speech is highly controlled by *top-down* attention, but that the signatures of *top-down* attentional control depend on the current acoustic adversity (i.e., SNR; see section 3.1). Especially in the most adverse listening condition, the late neural selective processing of the ignored talker plays a crucial role for the overall selective processing of the auditory scene. Importantly, this late neural processing of the ignored talker is accomplished by fronto-parietal than sensory brain regions, which highlights its involvement in *top-down* attentional control.

Study 2 showed, in contradiction to our hypothesis, that *alpha power* does not indicate the current demand for *top-down* attentional control in an auditory scene (see section 3.2). Further exploratory analysis did not reveal any relationship between *alpha power* modulation (or any other frequency band) and the acoustic adversity. A direct comparison of the attention-dependent *neural tracking* versus *alpha power* showed that during concurrent, continuous speech, neural signatures of auditory attention are predominantly emerging from the phase-locked neural responses (i.e, *neural tracking*).

Study 3 showed, in contradiction to our hypothesis, that neither the location of the talkers, the SNR, nor the interaction of location and SNR modulate *alpha power* (see section 3.3). Particularly, we showed that neither the modulation of *alpha power* at single EEG electrodes nor the whole-scalp hemispherical imbalance (i.e., *alpha power* lateralization) is indicative of the spatial focus of auditory attention in our task.

In studies 4–6, we investigated whether the feasibility of a reduced set of EEG electrodes including in-ear EEG detects a listener's focus of auditory attention, such that in-ear EEG might be used to inform a neurally steered hearing aid (see chapter 4).

Study 4 showed that within single subjects, spectrally resolved neural responses to stimuli rich of spectro-temporal modulation can be extracted from EEG electrodes placed around and inside the ear canal (see section 4.1). However, we also showed that an increased spectral resolution does not necessarily lead to a more precise prediction of the EEG signal.

Study 5 showed that single-channel in-ear EEG electrode configurations capture signatures of auditory selective attention to concurrent, continuous stimulus streams (see section 4.2). The attentional focus to dichotically presented tone streams and, more important, diotically presented concurrent speech could be detected within single subjects. The results of this study demonstrated the feasibility of single-channel EEG configurations attached to a hearing aid as a basis for the neural steering.

Study 6 showed that the listener's focus of attention can be detected under varying listening conditions with in-ear EEG electrode configurations (see section 4.3). Hence, we replicated study 5 by showing that single-channel electrode configuration detects a listener's focus of attention. Furthermore, under the most adverse condition, we found a similar increase of late *neural selectivity* in the response to the ignored as found before in study 1.

5.2 Attention-dependent *neural tracking* of speech: A consequence of spectro-temporal filtering?

We showed that the attention-modulated phase-locked neural responses to continuous speech in the form of *temporal response functions* show a high, replicable consistency across subjects (see sections 3.1, 4.2 & 4.3). We also showed that the modulation of those neural responses by *top-*

down attention emerges primarily in temporal, auditory cortical regions (see section 3.1). Due to the weak spatial resolution of EEG, we can neither directly infer on the exact anatomical position of this filter nor at which exact representation of the speech signal the filter is working. Source localizations of the attention-related difference pointed towards temporal and superior temporal regions. However, we have a comparably precise estimate of the time lag at which the filter affects the neural response, enabling us to infer on the anatomical location and its input based on existing literature.

The transformation from the pure acoustic representation of speech into a linguistic representation has been investigated in fMRI by means of comparing the activation by simple versus complex sounds or speech versus meaningless speech-like sounds (e.g., reversed speech, spectral rotation, vocoding). It was found that speech is preferentially processed in temporal lobe with increasing linguistic complexity being preferentially processed in the left superior temporal lobe along superior temporal gyrus and sulcus (Binder et al., 2000; Zatorre, 2002; Davis and Johnsrude, 2003; Narain et al., 2003; Hickok and Poeppel, 2004). Neuroanatomical models suggest that feedforward and feedback connections form a hierarchical structure where information is transmitted from core auditory cortical via belt towards parabelt areas (e.g., Sweet et al., 2005; Hackett et al., 2014). This highlights that the auditory cortex is branching the information about the incoming auditory signal and distributed this information to various brain areas. But at which stage does auditory attention come into play?

We observed earliest signatures of attention at around 80 ms at the transition between $P1_{TRF}$ and $N1_{TRF}$. Slow cortical components such as P1, N1 and P2 (or its neuromagnetic counterparts) were localized in *Heschl's gyrus* and *planum temporale*, which are core areas of the auditory cortex (Lütkenhöner and Steinsträter, 1998; Godey et al., 2001; Gascoyne et al., 2016). Based on those studies, we would assume that the *top-down* attentional filter we observed is working in a comparably early, spectro-temporal domain. This filter mechanism might be achieved by the attention-dependent tuning of auditory cortical neurons such that their spectro-temporal pattern of excitability (i.e., spectro-temporal receptive field) is biased towards the attended talker (Aertsen et. al, 1980; Escabi and Schreiner, 2002; Klein et al., 2000; Theunissen et al., 2000; Depireux et al., 2001; Fritz et al., 2007; Lakatos et al., 2013). However, there is evidence that linguistic articulatory features are already represented in comparably early (i.e., slow) cortical components such as the

P1 (i.e., P50m, Tavabi et al., 2007). It was also shown that a combination of spectro-temporal features and phonetic features best predict the slow cortical response to continuous speech (Di Liberto et al., 2015). Recently, it was shown that the transition between neural representations of spectro-temporal to linguistic features of continuous speech is achieved within the time frame of the N1 (Brodbeck et al., 2018), which shows that multiple neural encoding operations might be nested within one component of an ERP (or TRF) and are executed in parallel. More drastically, a recent study suggests that an intermediate linguistic representation does not even exist, since it is rendered redundant by an earlier representation merely consisting of acoustic edges (Daube et al., 2018). Since the task in our studies was to attend to either the female or male voice and participants were familiar with the voices after two blocks, we can assume that the attentional neural filters mainly worked in the spectro-temporal domain.

Importantly, the late component we found in the response to the ignored talker plays an exceptional role. This component was most prominent under the worst listening condition (see section 3.1). Since it appeared at comparably late time lags and showed a distinct parietal topography, we interpreted this component as a late signature of distractor suppression, which will be discussed below (see section 5.4).

One crucial question that arises at this point is whether the observed signatures of *top-down* attentional control are direct eyelet to the implementation of attentional filters or whether they are only the epiphenomenal consequence of a higher order *top-down* control. We tried to answer this question by the additional investigation of induced oscillations (see sections 3.2 & 3.3), where we specifically expected to find *alpha power* modulation to go along with the instantaneous demand for attentional control. Specifically, we expected to observe a relationship between *alpha power* modulation and the *neural tracking* response to speech. We did not find a clear evidence for such a higher order signature of *top-down* attentional control, as will be discussed in the next section.

5.3 *Alpha power* as a neural signature of *top-down* attentional control: What are we missing?

Our hypothesis about the attentional modulation of *alpha power* as a consequence of varying listening demand due to the SNR between – or spatial location of – the talkers was not confirmed

(see section 3.2 & 3.3). This is contradictory to previous studies which showed that *alpha power* is modulated by the current listening demand (e.g., Obleser and Weisz, 2012) and the focus of auditory spatial attention (Wöstmann et al., 2016). More generally, *alpha power* has been suggested to be involved in global *top-down* attentional processes by way of orchestrating sensory brain areas through inhibition (Klimesch e al., 2007) or channeling inputs along relevant neural pathways (Jensen and Mazaheri, 2010). We assume that our presented listening scenarios are more challenging compared to average real-life listening scenarios (Smeds et al., 2015). This leads to the question of why we did not observe a clear involvement of *alpha power* in the *top-down* selective neural processing of the concurrent talkers? This question will be discussed under two assumptions (Altman and Bland, 1995): First, modulation of *alpha power* is indeed absent, such that it is not involved in the neural selective processing of the talkers. Second, modulation of *alpha power* is present, but our design did not allow us to extract a consistent pattern.

5.3.1 Evidence for absence of attentional *alpha power* modulation

We designed our experiments to mirror real-world listening scenarios. That the *top-down* attentional modulation of *alpha power* was absent in these experiments might call into, the question arises, why conclusions made from earlier, trial-based studies: does *alpha power* modulation by *top-down* attention not hold for the more ecologically valid case of continuous listening?

One possible explanation is that selective listening does not necessarily involve alpha-induced *top-down* attentional control. Given that an earlier filter in primary sensory areas provides a representation of the attended talker which is sufficient to solve the current task, higher order filter strategies might not be exerted. Hence, the involvement of neural filter strategies related to *alpha power* might strongly depend on the actual listening task. As mentioned above, here the task instruction (i.e., attend female or male voice) might have led to the adaptation of spectro-temporal filters in auditory cortex, such that a *top-down* attentional adaptation of the filters was not necessary during listening but rather in the beginning of each block.

In contrast, trial-based task designs might reflect the alternating dis- and re-engagement into the task, which repeatedly triggers adaptation of neural filters and thus leads to a modulation of *alpha power*. Still, we expected to find SNR-dependent modulation of *alpha power* since neural filter strategies most likely vary between SNRs as the phase-locked neural responses indicated (see

section 3.1). A re-adaptation of neural filters that possibly induces an *alpha power* increase might be observed if participants were asked to switch their attentional focus within the blocks from one talker to the other (Mehraei et al., 2018). This might also interact with the current SNR, since a switch during an unbalanced SNR (i.e., ≠0 dB) would require a sudden adaptation of the neural filter strategies.

We assumed that a variation of the SNR within the range of −6 and +6 dB would significantly vary cognitive effort deployed to solve the listening task. This might have been not the case. Even if the SNR-variation led to the subjective impression of varying difficulty (as reported by the subjects), the deployed cognitive effort might not have varied. This would mean that the task did not explicitly trigger the deployment of more cognitive resources during worse SNRs. This point will be further discussed below (see section 0).

5.3.1 Absence of evidence for attentional *alpha power* modulation

Given that *alpha power* modulation reflects a ubiquitous neural mechanism involved in *top-down* neural processing of concurrent stimuli (Obleser et al., 2012; Obleser and Weisz, 2012; Wöstmann et al., 2015, 2016, 2017b), it may appear surprising that we have not observed a significant and consistent SNR- or location-dependent modulation (see section 3.2 & 3.3).

In our design, we varied the SNR stochastically. This means that the exact timing of an upcoming change as well as the rate of change was unpredictable (see section 3.2 & 3.3). Compared to the dominant temporal modulation spectrum of speech (see section 1.2), the SNR was low-frequent and sluggish. Thus, the adaptation to the instantaneous demand for attentional control might have not occurred under such a high temporal consistency as observed in trial-based designs, neither within nor across subjects. This might have led to the slightly increased *neural selectivity* observed in the alpha band, but non-consistent time-frequency response fields on the group level. Controversially, we also contrasted the power averaged across a whole period of a certain SNR (i.e., *plateau*), which should also capture dynamics that are not strictly locked to changes of the SNR. Still, an adaptation of neural filters might happen on an even larger temporal scale during listening to continuous speech.

Other studies indicate that the modulation of *alpha power* as a variable depending on external factors is following an inverted U-shape (instead of a linear relationship), which determines the

operational range of neural gain (Rajagovindan and Ding, 2010; Kloosterman et al., 2018). Since the adaptation of attentional filters involves the adjustment of neural gain (Willmore et al., 2014), modulation of *alpha power* might exist. However, since we always assumed a linear relationship between the external factor SNR (and location), we might not have captured such non-linear relationships.

5.4 Late response signature of reactive suppression of the ignored talker indicates increased listening effort

We argued that the signatures of auditory attention we observed in the phase-locked neural response to speech are mainly brought about by attentional filters tuned to the attended talker, such that the ignored talker gets filtered out in auditory cortex (see section 5.2; see also Mesgarani and Chang, 2012). However, under most adverse listening conditions, we additionally observed a late component in the response to the ignored talker, which unexpectedly localized to fronto-parietal brain regions. This component could not be explained by *bottom-up* attentional capture, but rather reflected *top-down* attentional control. We argue here that this component reflects the reactive suppression of the ignored talker, which is a mandatory effort in order to avoid the *bottom-up* capture of attention by the salient (i.e., *dominant*), to-be-ignored talker.

The suppression of salient, irrelevant stimuli that have the potential to capture attention in a *bottom-up* manner have been extensively studied in the visual modality (for a review see Gaspelin and Luck, 2019). Importantly, the potential of a stimulus to capture attention depends not only on its saliency, but also on its predictability, decision history and the overall *top-down* goals (Sawaki and Luck, 2010). This led to the formulation of the *signal suppression hypothesis* (for a review see Gaspelin and Luck, 2018), which states that salient, irrelevant stimuli are actively suppressed before they can capture attention. This view proposes an important interaction between *bottom-up* and *top-down* attention: The more precisely the *top-down* attentional goal is defined in terms of the to-be-attended and to-be-ignored stimulus features, the lower the potential that a to-be-ignored stimulus captures attention.

In recent studies, yet another dichotomous concept of attentional filtering has been used to explain the results: While proactive suppression refers to the pre-tuning of attentional filters based on to-be-expected features of the upcoming stimulus (e.g., Bonnefond and Jensen, 2012), reactive

suppression refers to the suppression of irrelevant stimuli after they have been encoded and were identified as to-be-ignored (e.g., Fukuda and Vogel, 2009). This means that, up to a certain stage, the ignored stimulus must be encoded in parallel to the attended stimulus in order to extract features crucial for suppression at a later stage. Importantly, reactive suppression must be achieved before the ignored stimulus captures *bottom-up* attention.

In our studies, participants were asked to attend to the male or the female voice before they self-paced the presentation by a button press, such that their attentional filters could have been proactively tuned (see section 5.2). Those filters might be optimized to let pass the features of the attended talker and to fend off the features of the ignored talker, as shown by the absence of an N1 and P2 component in the response to the ignored talker (Figure 5-1A). One interpretation of this attentional modulation is the proactive enhancement of all the features of the attended talker (Figure 5-1B, top). This might be an appropriate neural strategy if the signal of the attended talker is intact and not extensively degraded (i.e., masked) by the ignored talker, such as under positive SNRs. However, a second explanation is that the attentional filter is proactively suppressing the features of the ignored talker (Figure 5-1B, middle), such that only undegraded features of the attended talker pass the filter. This might be necessary when the attended talker is more degraded (i.e., masked) by the ignored talker. In any case, as soon as the incoming mixture passed the filter, missing features of the attended talker must be neurally restored (McDermott and Oxenham, 2008). Since speech signals are highly redundant (i.e., can be degraded and still be understood; Shannon et al., 1995) and predictable (e.g., Kutas et al., 2011), slight degradations can be easily compensated. However, more severe degradations result in increased demand for cognitive resources (Rabbit 1968, Pichora-Fuller and Singh, 2006).

A study that compared neural responses to clear and concurrent speech under active versus passive listening indicated that the neural filters are mainly suppressing the response to the ignored talker under selective attention (Kong et al., 2014). However, the authors asked the participants to passively listen to clear speech (while watching a silent movie), which might still have the potential to capture attention in a *bottom-up* manner if cognitive resources are not fully deployed to any other task (Lavie, 1995). Thus, the suppression of the ignored talker might strongly depend on the difficulty of the primary task, which modulates the demand for attentional control. There is behavioral evidence that the peripheral processing of ignored speech is reduced

under higher working-memory load caused by the primary task (Halin et al., 2015). Recently, it was also shown that *neural tracking* of attended speech as well as the 'automatic' stream segregation of ignored sounds is affected by working memory load (Hjortkjær et al., 2018; Molloy et al., 2018).

Importantly, during the most detrimental conditions, we found a late response to ignored talker (N2; see section 3.1). Since it facilitates neural selective processing, we interpret this component as a signature for active suppression of the ignored talker. But is this component a signature for proactive or reactive suppression? Since the N2 component appeared at a relatively late stage, we assume that it is a signature for reactive suppression of the ignored talker. This late neural processing of the ignored talker might be necessary because the earlier filter proactively tuned to the attended talker is not suffiecient to restore an intact representation of the attended talker (Figure 5-1B, bottom).

In contrast to all other components in temporal brain regions, we observed the late N2 component in fronto-parietal brain regions. The fronto-parietal network is associated with *top-down* attentional control of the dorsal auditory stream (e.g., Alain et al., 2001; Bidet-Caulet and Bertrand, 2005; for a review, see Hickok, 2012, Bizley and Cohen, 2013). Sustained frontal negativity has been linked to the gating of irrelevant information and its suppression (Chao and Knight, 1997), such that it is crucial for the maintenance of *top-down* attention. It was also shown that the lateral prefrontal cortex is involved in the facilitation of late auditory attention (Bidet-Caulet et al., 2015). In contrast to our findings, a negative component at prefrontal cortex was associated with facilitatory response to relevant stimuli (N_d component), whereas a positive, inhibitory, component was found in the response irrelevant stimuli (Alho et al., 1987; Michie et al., 1990; Alain et al., 1993). Note that in those ERP experiments, relevant and irrelevant stimuli were usually presented in some temporal order, such that the basis for attentional filtering was of temporal nature. In contrast, in our studies, concurrent speech was presented, which might reveal different components. Nevertheless, both the previous and our studies highlight that fronto-parietal brain regions are involved in the *top-down* attentional processing of the auditory input.

Given that the late N2 component is a signature of reactive suppression, it would imply that the ignored talker is encoded in parallel to the attended talker. Thus, more cognitive resources must be invested in order to work out a clean representation of the attended talker. Consequently,

the late N2 component might be an indicator for increased listening effort. Distinct frontal neural mechanisms for facilitation (i.e., enhancement of attended) and inhibitory (i.e., suppression of ignored) sounds have been shown to be differently affected by memory load (Bidet-Caulet et al., 2010). This highlights the intertwinement of *top-down* neural mechanisms and the working memory.

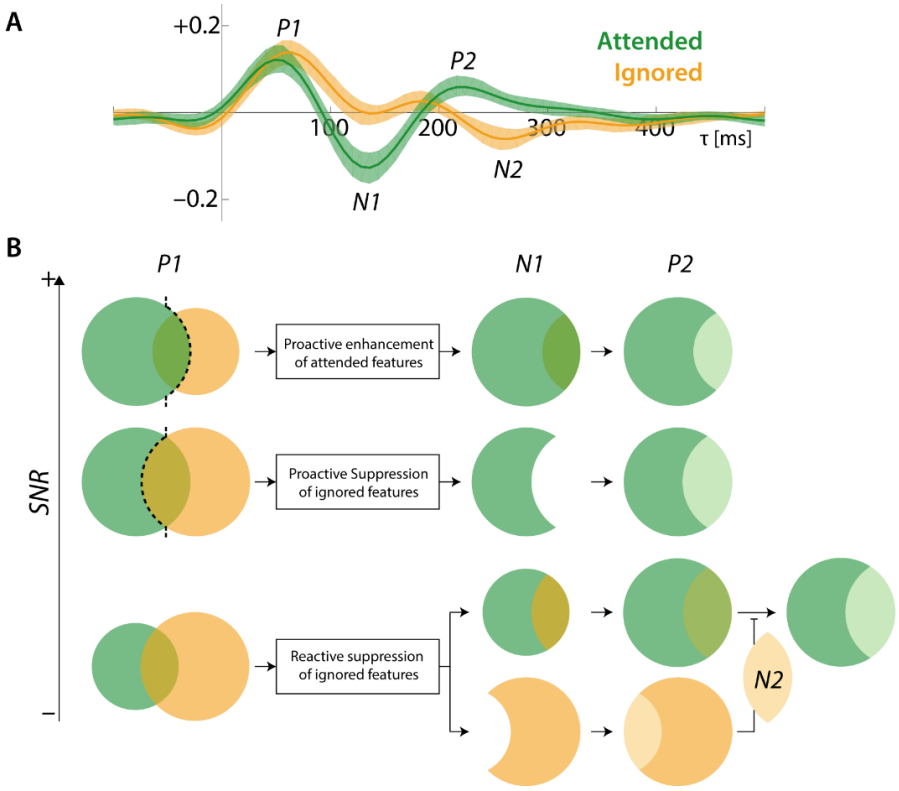

Figure 5-1: Neural filter strategies for selective attention. The overall goal of selective attention is to restore a clean representation of the attended stimulus (green dot) which is free from interference by ignored stimuli (orange dot). A) *Temporal response functions* for a *dominant* attended and a *dominant* ignored talker (adopted from study 1, see section 3.1). B) Neural filter strategies. Top: Proactive enhancement of the features of the attended talker. Middle: Proactive suppression of features of the ignored talker. Bottom: Reactive suppression of features of the ignored talker.

The interpretation of the N2 component as signature of late, reactive suppression was purely based on the electrophysiological outcome. Further studies should investigate the behavioral advantage (or disadvantage) related to this component. For example, considering that earlier, proactively tuned spectro-temporal filters cannot work as precise in hearing impaired subjects due to a degraded input, we would hypothesize that the late N2 component would appear already at better SNRs. This should be investigated by comparing hearing-impaired patients with matched controls. More fine-grained behavioral data should be recorded to evaluate the behavioral benefit of late, fronto-parietal reactive suppression.

5.5 The effort and risk of selective attentional filtering

In sum, the findings of our studies leave the neural basis for selective processing of concurrent speech open to some speculation. While we hypothesized that the *neural tracking* of speech and *alpha power* modulation are ubiquitous signatures of auditory attention, we found that only *neural tracking* is reliably indicating the attentional focus. Importantly, our analysis revealed late, inhibitory tracking of the ignored talker in fronto-parietal brain regions, a key functional role that has been previously assigned to *alpha power* (Klimesch et al., 2007). Hence, the phase-locked response to speech and the modulation of *alpha power* might be signatures of distinct neural strategies having a similar functional role. Since the behavioral relevance of *alpha power* for attentional filtering has been proven in previous studies (e.g., Obleser and Weisz, 2012; Wöstmann et al. 2016), we here provide a framework that aims at explaining why we did not observe clear evidence for the modulation of *alpha power*.

Two terms are key of the current framework: First, effort describes the cognitive resources that are deployed to solve the listening task (i.e., listening effort; e.g., Rönnberg et al., 2013). Risk describes the probability for making an error. Risk depends on how much effort is deployed. The more effort is deployed , the lower is the risk of making an error.

We argue that the deployment of certain neural strategies strongly depends on the task design, its demand, and the risk of error. In sum, these factors are assumed to be influential on the subject's motivation to solve a listening-in-noise task. This can be seen from a behavioral economics and neuroeconomics perspective (Kouneiher et al., 2009; Eckert et al., 2016).

The following framework is related to the *selective engagement hypothesis* (Hess, 2014; Hess, 2006), which argues that the engagement (i.e., deployment of limited cognitive resources) strongly depends on the motivation, which in turn depends on the self-rated consequences (positive and negative). Hess (2014) argues that limited cognitive resources in older adults result in a more selective engagement compared to young adults. This means that high cognitive resources can be spent even by older adults. However, due to the limited long-term cognitive resources, they are only spent in some, subjectively relevant tasks, but not in others. Here we argue that, independent of age, similar mechanisms determine whether certain neural strategies are deployed or not.

The current framework is based on the following assumptions. First, we assume that the attention-dependent *neural tracking* of speech (i.e., phase-locked neural response) and attention-dependent modulation of *alpha power* are neural signatures of two distinct neural strategies for selective attention. If applicable, the attentional modulation of *neural tracking* in sensory areas is less effortful than the modulation of *alpha power*. Second, both strategies can work proactively or reactively (see section 5.4). Third, the cognitive resources to process incoming information are limited (Kahneman, 1973; Pichora-Fuller et al., 2016) and this limitation is crucial for selective processing (Lavie, 1995). Fourth, the amount of invested cognitive resources (i.e., effort) depends on the subject's motivation, which in turn depends on the self-rated relevance to solve the task or, in other words, the willingness to lower the probability of making an error (here: risk; e.g., Eckert et al., 2016; Hess, 2014).

The general assumption is that the deployment of more cognitive effort reduces the risk of making an error (Figure 5-2, top). Minimally deployed effort results in chance performance, whereas the maximum of available cognitive resources is subject-specific and determines the degree to which the risk can be reduced.

We do not assume a linear relationship between *effort* and *risk.* Instead, we argue that the deployment of some effort lowers the risk (Δ risk) more than adding yet another unit of effort (Δ risk), which results in a risk function asymptotically approaching zero for towards higher effort. The higher the task demand, the higher the overall risk of making an error. Somewhere between minimal and maximal effort, the motivation of the subject determines, how much effort is deployed to reduce the risk down to a risk that is an adequate level. Importantly, the effort determines which neural strategies are used to solve the task (Figure 5-2, middle). Given that lower-level, sensory neural selective strategies are more efficient (but not always sufficient) than higher-level neural strategies, we assume that the attention-dependent *neural tracking* is generally used before *alpha power* modulation. Similarly, if proactive strategies are applicable, they are used before reactive strategies.

Here, we argue here that the audiobooks presented in our studies did not lead to a suffiently large motivation that subjects would invest all available cognitive resources. In our task, subjects only had to answer four questions at the end of a five-minute block. The risk of missing a few words in the to-be-attended stream due to interference of the ignored talker might not have

motivated the subjects to invest more cognitive effort into neural strategies related to *alpha power*. To the contrary, in tasks where *alpha power* modulation was observed, such as in (auditory) Posner-tasks (e.g., Wöstmann et al., 2016), subjects are asked to give a much more detailed response about the content of the stimulus. Effectively, this results in an increase of task demands (see Figure 5-2A). Consequently, subjects must deploy more effort in order to arrive at the same risk, which they do by using additional neural strategies, such as *alpha power* modulation.

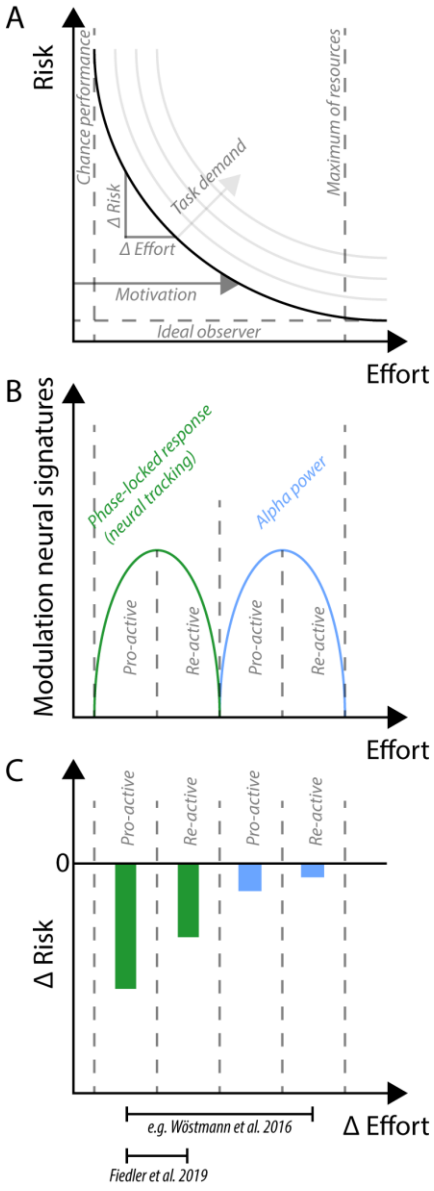

Figure 5-2: The effort and risk of attentional filtering. A) The investment of more cognitive effort reduces the risk of making an error. Motivation determines how much cognitive effort will be invested and is bound between minimal effort, resulting in chance performance, and maximal effort, which is limited by cognitive resources. B) Proposed neural signatures that reflect two different neural strategies: neural phase-locking (tracking) and *alpha power* modulation. If applicable, phase-locking in sensory areas is used before *alpha power* modulation in higher order areas. If applicable, pro-active filtering is used before re-active filtering. C) The reduction of risk (Δ risk) per additional unit of effort decreases towards higher order neural strategies.

5.6 Implications on neurally steered hearing aids

In the second part of this thesis (see chapter 4), we investigated the feasibility of in-ear EEG to record neural signatures of attention. Here, we primarily investigated the attentional modulation of phase-locked responses (i.e., *neural tracking* and *neural selectivity*), since we found it to be more predictive of a listener's attentional focus than modulation of *alpha power* (see section 3.1 & 3.2). We hypothesized that the focus of auditory attention can be detected based on a reduced set of electrodes such that a hearing aid could be provided with this information. We showed that a listener's focus of attention can be detected with a configuration consisting of only two electrodes, one in the ear canal and another next to the ear, respectively (see section 4.2). We replicated this finding and additionally showed that the detection of the attentional focus can also be achieved under varying listening conditions (see section 4.3).

To date, one issue regarding the application of EEG in hearing aids has been widely neglected so far: The detection of the focus of auditory attention is usually achieved based on the clean sound source signals, postulating that hearing aids are capable of a computational auditory scene analysis that returns clean source signals (e.g., Aroudi et al., 2016). Various hearing-aid-compatible algorithms for sound source separation exist (e.g., Wang and Brown, 2006). Each algorithm comes with strength and weaknesses, such that the estimated source signals might still be degraded under some circumstances. Hence, a decrease in the detection accuracy is expected.

Only a few studies have benchmarked the detection accuracy after the sound source signal was estimated by an ahead-slotted source signal separation. One approach relied on a neural network trained to separate voices based on a monaural signal (O'Sullivan et al., 2017). Another approach used an algorithm adapted from ICA to directly extract the broad-band temporal modulation of the sources based on the signals recorded on six hearing aid microphones (Van Eyndhoven et al., 2017). The latter has the advantage that the temporal modulation is rather low-frequent, so that its computational effort is low due to a reduced sampling rate. However, this algorithm can fail if background noise is diffuse and non-modulated. In this regard, the neural network approach can have an advantage, as it is directly trained to extract speech signals.

The application of linear models allowed us to inspect the neural response in the form of *temporal response functions* (TRFs). To do so, we first non-linearly transformed the sound waves into some representations that complied with our (and others; see section 2.3) assumption of how

sound is represented at the stage of the (auditory) cortex. Consequently, we could train linear models that were predictive of a listener's focus of attention. However, the overall explained variance in the EEG was quite small (i.e., less than 1%). Eventhough we do not know the to-be-explained variance, there is enough headroom left which could be possibly filled to some degree by fitting non-linear kernels (e.g., artificial neural networks). For instance, convolutional neural networks have been shown to improve the detection of the attended talker (Deckers et al., 2018). Scientifically, artificial neural networks may have the disadvantage that once fitted, the underlying computations are hard to interpret (e.g., Kell et al., 2018). However, for neurally steered hearing aids, artificial neural networks have the potential to significantly improve detection accuracy.

Our models were not only linear, they were also time-invariant. This might have restricted the explained variance as well. For example, it is known that the repeated presentation of a stimulus leads to a suppression of the neural response (e.g., Nagy and Rugg, 1989). Hence, the stimulus history plays an important role. We also showed that the morphology (i.e., amplitude and latency) of the phase-locked neural responses varied across SNRs, such that a variable response model might improve the detection of the attended talker (and the ignored) as well.

The high replicability of attended-talker detection from EEG should further encourage researchers to move experiments out of the lab and to prototype first neurally steered hearing aids in real life (O'Sullivan et al., 2014, 2015; Mirkovic et al. 2015, 2016; Biesmans et al., 2016; Das et al., 2016, 2018; Fiedler et al., 2017, 2019; Fuglsang et al., 2017). The manifold of listening scenarios cannot be simulated in the laboratory, such that some issues might be only detected in real-life listening scenarios.

For example, in the laboratory, we assume that a hearing-aid user always tries to listen attentively, however in real-life, they might just attend to some visual object, ignoring auditory input. Hence, a hearing aid should be able to detect if the user is attending to any sound source in the first place, before it detects *which* source is attended. For this purpose, parieto-occipital *alpha power* modulation might be indicative, since it is enhanced during auditory compared to visual attention (Adrian, 1944; Fu et al., 2001).

A factor that may positively influence detection accuracy in the long run is the interaction between reward (e.g., the hearing aid increases signal-to-noise ratio and listening gets easier) and

the neural signatures of attention (i.e., neurofeedback; Zink et al., 2017). In theory, hearing aid users may adapt a certain way of attending as soon as they experience that the hearing aid is correctly detecting the attended talker. In turn, this may lead to more distinct attentional neural signatures. Such an interaction should be ideally investigated in a longitudinal study. It can be debated if the listening success (i.e., understanding a talker) serves as a sufficient reward to induce such mechanisms, or if additional reward is needed.

Even if we showed that the *neural tracking* is most indicative of a listener's focus of attention, additional signatures of auditory attention during real-life listening scenarios might be discovered. For example, the lack of attention-indicative *alpha power* modulation might have been due to the one-way, non-interactive task design. If hearing aid users engage into real-life conversations, *alpha power* might also reveal valuable signatures of the attentional state. Furthermore, there might be other signatures in the EEG that are not directly related to the selective processing *per se*, but rather appear as artifacts caused by effortful listening. For example, a listener may be frowning if the listening conditions are too bad (e.g., due to low SNR), which results in muscle artifacts. What is a to-be-avoided confound for a neuroscientist, might be a valuable signature for an individually trained classifier. Hence, during the training of a classifier, the EEG signal should not be restricted to the signatures known to be indicative of attention *per se*.

In sum, we conclude that the *neural tracking* of speech is valuable signature for the steering of a hearing aid that can be captured by only a few electrodes attached to a hearing aid. However, the manifold of listening scenarios might reveal the full range of neural signatures related to selective attention. Prototypes of neurally steered hearing aids should be tested in real-life listening scenarios in order to evaluate the full potential and to detect possible challenges.

5.7 Limitations of the present research

5.7.1 Insufficient behavioral data

In the current studies, we mainly presented continuous speech of several minutes in order to infer the neural dynamics of listening in realistic scenarios. After each block, subjects were asked to answer several questions about the content of the to-be-attended story. Unsurprisingly, we could not find any relationship between number of correctly answered questions and the *neural*

tracking of speech. However, the behavioral relevance of the differential *neural tracking* between attended and ignored speech should be of interest. Especially, to what extent the late *neural tracking* of the ignored talker (see section 3.1) affects behavior should be investigated. Follwoinf our interpretation that this component avoids *bottom-up* attentional capture, the absence of this component should result in stream confusions.

Particularly, it is of interest to what extent the *neural tracking* of speech is related to speech intelligibility and comprehension. The reconstruction of attended speech in noise from MEG has been shown to correlate with the self-rated intelligibility across subjects (Ding and Simon, 2013). While a neural measure of instantaneous speech comprehension is still missing, some studies showed that speech comprehension is related to the reconstruction accuracy (Vanthornhout et al., 2018; Verschueren et al., 2018). However, to achieve variability in speech intelligibility, the stimulus itself must be acoustically manipulated (e.g., vocoding or background noise). Thus, a lowered *neural tracking* of speech might be due to the acoustic degradation. Acoustic degradation might also affect speech comprehension. However, to directly test the influence of *neural tracking* on speech comprehension, the mediation from the acoustics via *neural tracking* to the behavioral outcome must be tested in a single-trial fashion.

The low number of questions asked in our studies did not allow to sample behavior at a high rate. Furthermore, some questions were formulated in a rather general fashion and did not precisely point to a certain sentence or word of the story. To the contrary, in trial-based designs, much more fine-grained behavioral data can be recorded, allowing for the investigation of the behavioral relevance of neural measures. On the other hand, trial-based designs are lacking the degree of ecological validity that comes with continuous stimulation. It is therefore important to gather more behavioral data from continuous stimulation.

One possible is to randomly stop the presentation and ask subjects to repeat the last sentence (O'Sullivan et al., 2017). This has the advantage of introducing a call-and-response structure, similar to a real conversation. It would also allow for the investigation of a hazard-dependent *neural tracking* of speech: the longer the sequence, the more likely the presentation will stop. Hence, subjects may invest more cognitive resources in the processing towards the expected end of the block, such that *neural tracking* might become a function of the implicit hazard (Herbst et al., 2018). This might also be accompanied by modulation of induced oscillations.

Another possible approach to gather more behavioral data from continuous stimulation is a memory task. After subjects listened to concurrent, continuous speech, they were invited later (e.g., the next day) and listen to random probes (e.g., five seconds) of both the to-be-attended and to-be-ignored talkers. Subjects should be asked to indicate, whether they have heard the probe before. Additionally, some random probes should be presented to infer the rate of guessing. Generally, we would expect that more of the to-be-attended probes are remembered. Importantly, remembered probes of the to-be-ignored talker indicate stream confusions. Consequently, the *neural tracking* of speech and induced oscillations can be investigated based on the behavioral responses.

In sum, the lack of behavioral data prevented us from estimating whether listening performance depends on the *neural tracking* of speech, even though differential *neural tracking* between attended and ignored speech might indicate it. However, as shown above, there are ways to present continuous speech and still collect more fine-grained behavioral data.

5.7.2 Ecological validity

We claimed that the presentation of continuous speech matches real-world listening scenarios more closely than trial-based designs. There are several arguments in favor of continuous stimulation (Hamilton and Huth, 2018; Alexandrou et al., 2018). However, there are also some aspects that still hinder conclusions on real listening.

First, we presented some pre-selected audiobooks, which might have an impact on the subject's motivation. In real environments, listeners can decide whom to listen to and this decision might be based on the listener's goal. As discussed above (see section 0), the motivation might interact with the deployed effort and consequently with the involved neural strategies. This might explain why we did not observe an involvement of *alpha power* modulation in selective attentional filtering (see section 3.2 & 3.3). This can be circumvented (at least to some degree) by offering several stories to choose from.

Second, listening to speech in noise is usually happening in interactive scenarios (e.g., in a bar), while listening to the same voice for a couple of minutes is rather happening in quiet (e.g., in a lecture). The constant, unidirectional flow of information might have put subjects into a state of more general listening, meaning that they were only interested in the general content rather than

in every single word. As discussed above (see section 0), this might have also an effect on the involved neural strategies, since the participants deploy effort more constantly instead of instantaneously. Studying the neural mechanisms in real-world conversations needs comprehensive planning and post-processing, but it is increasingly established in recent research (e.g., Bevilacqua et al., 2018; Poulsen et al., 2017). Contrary to our initial assumption, a trial-based design might even more closely reflect those interactions in terms of bidirectional flow of information and instantaneously deployed effort.

Third, listening to a talker in noise usually is accompanied by visual information. Especially lip-reading is known to facilitate speech comprehension in noise (e.g., Macleod and Summerfield, 1987). Thus, effects of cross-modal integration are largely neglected with our design. For example, the cross-modal integration of continuous audio-visual speech has been shown to reveal greater reconstruction accuracies than the sum of its parts (Crosse et al., 2015). Regarding neurally steered hearing aids, this might be another source of information we have not investigated with our design.

Fourth, not only the presented signal but also the noise is important. Here we always presented a to-be-ignored talker of opposite gender. First, this allowed us to leave out the factor of spatial segregation and still provide an unambiguous cue to the participants. Second, it allowed us to investigate the neural processing of ignored speech. However, in real listening scenarios, noise is much more variable (both spatially and spectro-temporally) and sometimes it is more than one sound source that must be ignored.

In sum, with our designs, we studied the neural mechanisms underlying the selective processing of continuous attended and ignored speech. However, auditory selective attention has to be investigated in further studies, thoughtfully designed to capture different aspects of real-world behavior.

5.8 Conclusions

This thesis aimed to explain the modulation of the two neural signatures of auditory attention: *neural tracking* and *alpha power*. The core results of this thesis suggest that the two neural signatures reflect two different neural strategies. While *neural tracking* is most prominent in sensory areas (but also emerges from fronto-parietal brain regions), *alpha power* was not following the expected pattern.

We conclude that neural attentional filters can work at earlier or later representations along the hierarchical processing of the (sensory) pathway. At which stage selection is realized strongly depends on the current conditions. In our tasks, attentional filtering is mainly achieved at a comparably early, spectro-temporal representation of the concurrent inputs, since their distinct features allow a dissociation at this stage. Spectro-temporal filters are proactively tuned to pass features of the to-be-attended voice and/or block features of the ignored talker. Only when listening conditions get more adverse, an additional component suppresses the ignored talker at a later a stage to avoid *bottom-up* attentional capture. Hence, neural filter strategies are highly adaptive and depend on the current condition.

We argue that the task itself strongly influences whether *alpha power* modulation can be observed. We suggest that our continuous listening tasks are substantially different from previous, trial-based designs. The differences explain the (non-)involvement of neural strategies that are related to *alpha power* in attentional selection. First, given that earlier, proactively tuned spectro-temporal filters provide a sufficiently clean representation of the attended talker (see above), the neural strategy that is mirrored by *alpha power* was not deployed in the current tasks. Second, the task demand and the risk of making an error did not make *alpha power* neural strategies obligatory. This means that presenting the same stimuli, but challenging subjects with a more demanding task would lead to the involvement of neural strategies related to *alpha power*.

Another scope of this thesis was to detect neural signatures of auditory attention at a reduced set of EEG electrodes, that can be attached to a hearing aid. Based on our results, we conclude that the *neural tracking* of speech provides the most reliable basis to inform a hearing aid about a listener's focus of attention. Even though continuous stimulation more closely emulates a real-world scenario (compared to trial-based designs), neurally steered hearing aids should be tested under even more natural conditions. This includes more interactive design such as conversations.

Especially the role of real-time neurofeedback should be investigated more in detail. Based on our earlier conclusions, we do not disqualify *alpha power* from providing valuable information for a hearing. To the contrary, our experiments might not have mirrored the operational range of *alpha power*. Especially the more global, cross-modal distribution of cognitive resources might reflect in *alpha power*, which would provide valuable information to steer a hearing aid.

In sum, this thesis provides a broad insight into the neural mechanisms underlying auditory selective attention under dynamically varying listening conditions. We provide strong evidence for *neural tracking* to reflect neural strategy that is highly adaptive to the current listening condition. In contrast, the role of *alpha power* in auditory selective attention remains elusive. It is up to future research to close the interpretational gap between continuous and trial-based designs.

References

Adrian ED (1944) Brain Rhythms*. Nature 153:360.

Aertsen AMHJ, Johannesma PIM, Hermes DJ (1980) Spectro-temporal receptive fields of auditory neurons in the grassfrog. Biol Cybern 38:235–248.

Alain C, Achim A, Richer F (1993) Perceptual context and the selective attention effect on auditory event-related brain potentials. Psychophysiology 30:572–580.

Alain C, Arnott SR, Picton TW (2001) Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. J Exp Psychol Hum Percept Perform 27:1072–1089.

Alexandrou AM, Saarinen T, Kujala J, Salmelin R (2018) Cortical entrainment: what we can learn from studying naturalistic speech perception. Lang Cogn Neurosci:1–13.

Alho K, Donauer N, Paavilainen P, Reinikainen K, Sams M, Näätänen R (1987) Stimulus selection during auditory spatial attention as expressed by event-related potentials. Biol Psychol 24:153–162.

Altman DG, Bland JM (1995) Absence of evidence is not evidence of absence. BMJ 311:485.

Aroudi A, Mirkovic B, Vos M De, Doclo S (2016) Auditory attention decoding with EEG recordings using noisy acoustic reference signals. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 694–698.

Bednar A, Lalor EC (2018) Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. Neuroimage 181:683–691.

Bentler RA (2005) Effectiveness of Directional Microphones and Noise Reduction Schemes in Hearing Aids: A Systematic Review of the Evidence. J Am Acad Audiol 16:473–484.

Bentler RA, Palmer C, Dittberner AB (2004) Hearing-in-Noise: Comparison of Listeners with Normal and (Aided) Impaired Hearing. J Am Acad Audiol 15:216–225.

Berger H (1929) Über das Elektrenkephalogramm des Menschen. Arch Psychiatr Nervenkr 87:527–570.

Berger H (1932) Über das Elektrenkephalogramm des Menschen. Arch Psychiatr Nervenkr 97:6–26.

Bevilacqua D, Davidesco I, Wan L, Oostrik M, Chaloner K, Rowland J, Ding M, Poeppel D, Dikker S (2018) Brain-to-Brain Synchrony and Learning Outcomes Vary by Student–Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. J Cogn Neurosci:1–11.

Bidet-Caulet A, Bertrand O (2005) Dynamics of a temporo-fronto-parietal network during sustained spatial or spectral auditory processing. J Cogn Neurosci 17:1691–1703.

Bidet-Caulet A, Buchanan KG, Viswanath H, Black J, Scabini D, Bonnet-Brilhault F, Knight RT (2015) Impaired Facilitatory Mechanisms of Auditory Attention After Damage of the Lateral Prefrontal Cortex. Cereb Cortex 25:4126–4134.

Bidet-Caulet A, Mikyska C, Knight RT (2010) Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. Neuroimage 50:277–284.

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} Trans Neural Syst Rehabil Eng 25:402–412.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000) Human Temporal Lobe Activation by Speech and Nonspeech Sounds. Cereb Cortex 10:512–528.

Bizley JK, Cohen YE (2013) The what, where and how of auditory-object perception. Nat Publ Gr 14:693–707.

Bleichner MG, Debener S (2017) Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. Front Hum Neurosci 11:163.

Bleichner MG, Lundbeck M, Selisky M, Minow F, Jager M, Emkes R, Debener S, De Vos M (2015) Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? Physiol Rep 3:e12362.

Bleichner MG, Mirkovic B, Debener S (2016) Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. J Neural Eng 13:1–13.

Bonnefond M, Jensen O (2012) Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. Curr Biol 22:1969–1974.

Bosman CA, Womelsdorf T, Desimone R, Fries P (2009) A Microsaccadic Rhythm Modulates Gamma-Band Synchronization and Behavior. J Neurosci 29:9471–9480.

Bregman AS (1990) Auditory Scene Analysis. MIT Press.

Broadbent DE (1958) Perception and Communication. Oxford: Pergamon Press.

Brodbeck C, Hong LE, Simon JZ (2018) Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Curr Biol 28:3976–3983.

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Edmund C (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr Biol 28:803–809.

Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? J Neurosci Methods 137:321–332.

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

Chait M, Cheveigné A De, Poeppel D, Simon JZ (2010) Neural dynamics of attending and ignoring in human auditory cortex. Neuropsychologia 48:3262–3271.

Chao LL, Knight RT (1997) Prefrontal deficits in attention and inhibitory control with aging. Cereb cortex 7:63–69.

Cherry EC (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am 25:975–979.

Chi T, Ru P, Shamma SA (2005) Multiresolution Spectrotemporal Analysis of Complex Sounds. J Acoust Soc Am 118:887–906.

Cohen MX, Gulbinaite R (2014) Five methodological challenges in cognitive electrophysiology. Neuroimage 85:702–710.

Cokely JA, Hall JW (1991) Frequency resolution for diotic and dichotic listening conditions compared using the bandlimiting measure and a modified bandlimiting measure. J Acoust Soc Am 89:1331–1339.

Combrisson E, Jerbi K (2015) Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods 250:126–136.

Comon P (1994) Independent component analysis, A new concept? Signal Processing 36:287–314.

Connor CE, Egeth HE, Yantis S (2004) Visual Attention: Bottom-Up Versus Top-Down. Curr Biol 14:850–852.

Crosse MJ, Butler JS, Lalor EC (2015) Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. J Neurosci 35:14195–14204.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci 10:604.

Das N, Biesmans W, Bertrand A, Francart T (2016) The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. J Neural Eng 13:056014.

Daube C, Ince R, Gross J (2018) Phoneme-level processing in low-frequency cortical responses to speech explained by acoustic features. bioRxiv:https://doi.org/10.1101/448134.

David O, Kilner JM, Friston KJ (2006) Mechanisms of evoked and induced responses in MEG/EEG. Neuroimage 31:1580–1591.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Debener S, Emkes R, De Vos M, Bleichner MG (2015) Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. Sci Rep 5:16743.

Deckers L, Das N, Hossein Ansari A, Bertrand A, Francart T (2018) EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks. bioRxiv:https://doi.org/10.1101/475673.

Denk F, Grzybowski M, Ernst SMA, Kollmeier B, Debener S, Bleichner MG (2018) Event-Related Potentials Measured From In and Around the Ear Electrodes Integrated in a Live Hearing Device for Monitoring Sound Perception. Trends Hear 22:1–14.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex. J Neurophysiol 85:1220–1234.

Deutsch JA, Deutsch D (1963) Attention: some theoretical considerations. Psychol Rev 70:80–90.

Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr Biol 25:2457–2465.

Ding N, Melloni L, Yang A, Wang Y, Zhang W (2017) Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). Front Hum Neurosci 11:1–9.

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

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.

Doclo S, Gannot S, Moonen M, Spriet A (2010) Acoustic Beamforming for Hearing Aid Applications. In: Handbook on Array Processing and Sensor Networks, pp 269–302. John Wiley & Sons, Ltd.

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85 Pt 2:761–768.

Duncan J (2006) EPS mid-career award 2004: Brain mechanisms of attention. Q J Exp Psychol 59:2–27.

Dyson BJ, Alain C, He Y (2005) Effects of visual attentional load on low-level auditory scene analysis. Cogn Affect Behav Neurosci 5:319–338.

Eckert MA, Teubner-Rhodes S, Vaden Jr KI (2016) Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. Ear Hear 37 Suppl 1:101S–10S.

Efron B (1979) Bootstrap methods: Another look at the jackknife. Ann Stat 7:1–26.

Egeth HE, Yantis S (1997) Visual Attention: Control, Representation, and Time Course. Annu Rev Psychol 48:269–297.

Elhilali M, Xiang J, Shamma SA, Simon JZ (2009) Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol 7:e1000129–e1000129.

Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. J Neurosci 22:4114–4131.

Etard O, Kegler M, Braiman C, Forte AE, Reichenbach T (2018) Real-time decoding of selective attention from the human auditory brainstem response to continuous speech. bioRxiv:https://doi.org/10.1101/259853.

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

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

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

Forte AE, Etard O, Reichenbach T (2017) The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife 6:e27203:1–12.

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

Frey JN, Mainy N, Lachaux J-P, Müller N, Bertrand O, Weisz N (2014) Selective Modulation of Auditory Cortical Alpha Activity in an Audiovisual Spatial Attention Task. J Neurosci 34:6634–6639.

Fritz JB, Elhilali M, David S V, Shamma SA (2007) Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? Hear Res 229:186–203.

Fu K-MG, Foxe JJ, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2001) Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto–occipital alpha-band oscillations. Cogn Brain Res 12:145–152.

Fuglsang SA, Dau T, Hjortkjær J (2017) Noise-robust cortical tracking of attended speech in real-world acoustic scenes. Neuroimage 156:435–444.

Fukuda K, Vogel EK (2009) Human Variation in Overriding Attentional Capture. J Neurosci 29:8726–8733.

Gascoyne L, Furlong PL, Hillebrand A, Worthen SF, Witton C (2016) Localising the auditory N1m with event-related beamformers: localisation accuracy following bilateral and unilateral stimulation. Sci Rep 6:31052.

Gaspelin N, Luck SJ (2018) The Role of Inhibition in Avoiding Distraction by Salient Stimuli. Trends Cogn Sci 22:79–92.

Gaspelin N, Luck SJ (2019) Inhibition as a potential resolution to the attentional capture debate. Curr Opin Psychol 29:12–18.

Giordano BL, Ince RAA, Gross J, Schyns PG, Panzeri S, Kayser C (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. eLife 6:1–27.

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liégeois-Chauvel C (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin Neurophysiol 112:1850–1859.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892.

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic Correlates of Streaming in Human Auditory Cortex. J Neurosci 25:5382–5388.

Hackett TA, de la Mothe LA, Camalier CR, Falchier A, Lakatos P, Kajikawa Y, Schroeder CE (2014) Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. Front Neurosci 8:72.

Haegens S, Händel BF, Jensen O (2011) Top-Down Controlled Alpha Band Activity in Somatosensory Areas Determines Behavioral Performance in a Discrimination Task. J Neurosci 31:5197–5204.

Halin N, Marsh JE, Sörqvist P (2015) Central load reduces peripheral processing: Evidence from incidental memory of background speech. Scand J Psychol 56:607–612.

Hambrook DA, Tata MS (2014) Theta-band phase tracking in the two-talker problem. Brain Lang 135:52–56.

Hamilton LS, Edwards E, Chang EF, Hamilton LS, Edwards E, Chang EF (2018) A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. Curr Biol 28:1860–1871.

Hamilton LS, Huth AG (2018) The revolution will not be controlled: natural stimuli in speech neuroscience. Lang Cogn Neurosci:https://doi.org/10.1080/23273798.2018.1499946.

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.

Hautus MJ, Johnson BW (2005) Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. J Acoust Soc Am 117:275–280.

Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. PNAS 109:20095–20100.

Herbst SK, Fiedler L, Obleser J (2018) Tracking temporal hazard in the human electroencephalogram using a forward encoding model. eNeuro 5.

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

Hess TM (2014) Selective Engagement of Cognitive Resources: Motivational Influences on Older Adults' Cognitive Functioning. Perspect Psychol Sci a J Assoc Psychol Sci 9:388–407.

Hickok G (2012) The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. J Commun Disord 45:393–402.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99.

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. Science 182:177–180.

Hjortkjær J, Märcher-Rørsted J, Fuglsang SA, Dau T (2018) Cortical oscillations and entrainment in speech processing during working memory load. Eur J Neurosci 0.

Hoerl AE, Kennard RW (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12:55–67.

Hohmann V (2002) Frequency analysis and synthesis using a Gammatone filterbank. Acta Acust united with Acust 88:433–442.

Holdgraf CR, Rieger JW, Micheli C, Martin S, Knight RT, Theunissen FE (2017) Encoding and Decoding Models in Cognitive Electrophysiology. Front Syst Neurosci 11:61.

Horton C, Srinivasan R, Zmura MD (2014) Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party.' J Neural Eng 11:046015.

Horton C, Zmura MD, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol 109:3082–3093.

Howard MF, Poeppel D (2010) Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. J Neurophysiol 104:2500–2511.

Huang N, Elhilali M (2017) Auditory salience using natural soundscapes. J Acoust Soc Am 141:2163–2176.

Ince RAA, Mazzoni A, Petersen RS, Panzeri S (2010) Open source tools for the information theoretic analysis of neural data. Front Neurosci 4:62–70.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Jia J, Liu L, Fang F, Luo H (2017) Sequential sampling of visual objects during sustained attention. PLOS Biol 15:1–19.

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

Karrasch M, Laine M, O Rinne J, Rapinoja P, Sinervä E, Krause CM (2006) Brain oscillatory responses to an auditory-verbal working memory task in mild cognitive impairment and Alzheimer's disease. Int J Psychophysiol 59:168–178.

Katsuki F, Constantinidis C (2013) Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. Neurosci 20:509–521.

Kaya EM, Elhilali M (2014) Investigating bottom-up auditory attention. Front Hum Neurosci 8:327.

Kaya EM, Elhilali M (2017) Modelling auditory attention. Phil Trans R Soc B 372:1–10.

Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. Curr Biol 15:1943–1947.

Kayser SJ, Ince RAA, Gross J, Kayser C (2015) Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. J Neurosci 35:14691–14701.

Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLOS Biol 16:1–19.

Keitel A, Ince RAA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 147:32–42.

Kell AJE, Yamins DLK, Shook EN, Norman-Haignere S V, McDermott JH (2018) A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. Neuron 98:630–644.e16.

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party." J Neurosci 30:620–628.

Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J Comput Neurosci 9:85–111.

Klem GH, Luders H, Jasper HH, Elger C (1999) The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl 52:3–6.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: The inhibition-timing hypothesis. Brain Res Rev 53:63–88.

Kloosterman NA, de Gee JW, Werkle-Bergner M, Lindenberger U, Garrett DD, Fahrenfort JJ (2018) Humans strategically shift decision bias by flexibly adjusting sensory evidence accumulation in visual cortex. bioRxiv:https://doi.org/10.1101/229989.

Kong Y-Y, Somarowthu A, Ding N (2015) Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech. J Assoc Res Otolaryngol 16:783–796.

Kong YY, Mullangi A, Ding N (2014) Differential modulation of auditory responses to attended and unattended speech in different listening conditions. Hear Res 316:73–81.

Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: A review of classification techniques. Emerg Artif Intell Appl Comput Eng 160:3–24.

Kouneiher F, Charron S, Koechlin E (2009) Motivation and cognitive control in the human prefrontal cortex. Nat Neurosci 12:939–945.

Kraus N, Nicol T (2008) Auditory evoked potentials. In: Encyclopedia of Neuroscience, pp 214–218. Springer.

Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27.

Kutas M, DeLong KA, Smith NJ (2011) A look around at what lies ahead: Prediction and predictability in language processing. In: Predictions in the brain: Using our past to generate a future (Bar M, ed). Oxford University Press.

Lachaux J, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring Phase Synchrony in Brain Signals. Hum Brain Mapp 8:194–208.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. Science 320:110–113.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

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

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ (2006) The VESPA: A method for the rapid estimation of a visual evoked potential. Neuroimage 32:1549–1561.

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

Lavie N (1995) Perceptual load as a necessary condition for selective attention. J Exp Psychol Hum Percept Perform 21:451–468.

Leiberg S, Lutzenberger W, Kaiser J (2006) Effects of memory load on cortical oscillatory activity during auditory pattern working memory. Brain Res 1120:131–140.

Levitt H (2001) Noise reduction in hearing aids: A review. J Rehabil Res Dev 38:111–122.

Levitt H, Rabiner LR (1967) Predicting Binaural Gain in Intelligibility and Release from Masking for Speech. J Acoust Soc Am 42:820–829.

Looney D, Kidmose P, Park C, Ungstrup M, Rank ML, Rosenkranz K, Mandic DP (2012) The In-the-Ear Recording Concept: User-Centered and Wearable Brain Monitoring. IEEE Pulse 3:32–42.

Lopes da Silva F (2013) EEG and MEG: Relevance to neuroscience. Neuron 80:1112–1128.

Lunner T, Gustafsson F (2016) Hearing device with brainwave dependent audio processing. U.S. Patent No. 9,432,777. 30 Aug. 2016.

Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol 8:e1000445.

Luo H, Poeppel D (2007) Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron 54:1001–1010.

Lütkenhöner B, Steinsträter O (1998) High-Precision Neuromagnetic Study of the Functional Organization of the Human Auditory Cortex. Audiol Neurotol 3:191–213.

Lyons RG (2004) Understanding Digital Signal Processing (2Nd Edition). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Macleod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception in noise. Br J Audiol 21:131–141.

Makeig S, Debener S, Onton J, Delorme A (2004) Mining event-related brain dynamics. Trends Cogn Sci 8:204–210.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

McDermott JH, Oxenham AJ (2008) Spectral completion of partially masked sounds. Proc Natl Acad Sci 105:5939–5944.

McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71:926–940.

McGarrigle R, Munro KJ, Dawes P, Stewart AJ, David R, Barry JG, Amitay S (2014) Listening effort and fatigue: What exactly are we measuring? Int J Audiol 53:433–445.

McMahon CM, Boisvert I, de Lissa P, Granger L, Ibrahim R, Lo CY, Miles K, Graham PL (2016) Monitoring Alpha Oscillations and Pupil Dilation across a Performance-Intensity Function. Front Psychol 7:745.

Mehraei G, Shinn-Cunningham B, Dau T (2018) Influence of talker discontinuity on cortical dynamics of auditory spatial attention. Neuroimage 179:548–556.

Melara RD, Rao A, Tong Y (2002) The duality of selection: Excitatory and inhibitory processes in auditory selective attention. J Exp Psychol 28:279–306.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E (2007) The role of auditory cortex in the formation of auditory streams. Hear Res 229:116–131.

Michie PT, Bearpark HM, Crawford JM, Glue LCT (1990) The nature of selective attention effects on auditory event-related potentials. Biol Psychol 30:219–250.

Mikkelsen KB, Kappel SL, Mandic DP, Kidmose P (2015) EEG Recorded from the Ear: Characterizing the Ear-EEG Method. Front Neurosci 9:438.

Mikkelsen KB, Kidmose P, Hansen LK (2017) On the Keyhole Hypothesis: High Mutual Information between Ear and Scalp EEG. Front Hum Neurosci 11:341.

Mirkovic B, Bleichner MG, De Vos M, Debener S (2016) Target speaker detection with concealed EEG around the ear. Front Neurosci 10:1–11.

Mirkovic B, Debener S, Jaeger M, Vos M De (2015) Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. J Neural Eng 12:046007.

Moher J, Egeth HE (2012) The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. Attention, Perception, Psychophys 74:1590–1605.

Molloy K, Lavie N, Chait M (2018) Auditory figure-ground segregation is impaired by high visual load. J Neurosci:https://doi.org/10.1423/JNEUROSCI.2518-18.2018.

Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol 70:909–919.

Näätänen R (1975) Selective attention and evoked potentials inhumans — A critical review. Biol Psychol 2:237–307.

Näätänen R, Picton TW (1987) The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. Psychophysiology 24:375–425.

Nagy ME, Rugg MD (1989) Modulation of Event-Related Potentials by Word Repetition: The Effects of Inter-Item Lag. Psychophysiology 26:431–436.

Narain C, Scott S, J S Wise R, Rosen S, Leff A, Iversen SD, Matthews P (2004) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362–1368.

Narayanan AM, Bertrand A (2018) The effect of miniaturization and galvanic separation of EEG sensor devices in an auditory attention detection task. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 77–80.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410.

O'Sullivan JA, Chen Z, Herrero J, McKhann GM, Sheth SA, Mehta AD, Mesgarani N (2017) Neural decoding of attentional selection in multi-speaker environments without access to clean sources. J Neural Eng 14:056001.

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex 25:1697–1706.

O'Sullivan JA, Reilly RB, Lalor EC (2015) Improved decoding of attentional selection in a cocktail party environment with EEG via automatic selection of relevant independent components. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS 2015–Novem:5740–5743.

Obleser J (2014) Putting the Listening Brain in Context. Linguist Lang Compass 8:646–658.

Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn Sci 13:14–19.

Obleser J, Weisz N (2012) Suppressed Alpha Oscillations Predict Intelligibility of Speech and its Acoustic Details. Cereb Cortex 22:2466–2477.

Obleser J, Wöstmann M, Hellbernd N, Wilsch A, Maess B (2012) Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network. J Neurosci 32:12376–12383.

Olguin A, Bekinschtein TA, Bozic M (2018) Neural Encoding of Attended Continuous Speech under Different Types of Interference. J Cogn Neurosci 30:1606–1619.

Oostenveld R, Fries P, Maris E, Schoffelen J (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci 2011.

Patterson RD (1976) Auditory filter shapes derived with noise stimuli. J Acoust Soc Am 59:640–654.

Peelle JE (2018) Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. Ear Hear 39:204–214.

Petersen EB, Wöstmann M, Obleser J, Lunner T (2016) Neural tracking of attended versus ignored speech is differentially affected by hearing loss. J Neurophysiol 117:18–27.

Petersen EB, Wöstmann M, Obleser J, Stenfelt S, Lunner T (2015) Hearing loss impacts neural alpha oscillations under adverse listening conditions. Front Psychol 6:1–11.

Pfurtscheller G (2003) Induced Oscillations in the Alpha Band: Functional Meaning. Epilepsia 44:2–8.

Pfurtscheller G, Aranibar A (1977) Event related desynchronization detected by power measurement of scalp EEG. Electroencephalogr Clin Neurophysiol 42:817–826.

Pfurtscheller G, Klimesch W (1992) Functional Topography During a Visuoverbal Judgment Task Studied with Event-Related Desynchronization Mapping. J Clin Neurophysiol 9:120–131.

Pfurtscheller G, Stancák A, Neuper C (1996) Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. Int J Psychophysiol 24:39–46.

Pichora-Fuller KM, Kramer S, Eckert MA, Edwards B, Hornsby B, E Humes L, Lemke U, Lunner T, Matthen M, Mackersie C, Naylor G, Phillips N, Richter M, Rudner M, Sommers M, Tremblay K, Wingfield A (2016) Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). Ear Hear 37 Suppl 1:5S–27S.

Pichora-Fuller KM, Schneider BA, Daneman M (1995) How young and old listen to and remember speech in noise. J Acoust Soc Am 91:593–608.

Pichora-Fuller KM, Singh G (2006) Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. Trends Amplif 10:29–59.

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

Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ (1999) Intracerebral Sources of Human Auditory-Evoked Potentials. Audiol Neurotol 4:64–79.

Picton TW, Hillyard SA (1974) Human auditory evoked potentials. ii: effects of attention. Electroencephalogr Clin Neurophysiol 36:191–199.

Piquado T, Isaacowitz D, Wingfield A (2010) Pupillometry as a measure of cognitive effort in younger and older adults. Psychophysiology 47:560–569.

Pomper U, Chait M (2017) The impact of visual gaze direction on auditory object tracking. Sci Rep 7:1–16.

Popov V, Ostarek M, Tenison C (2018) Practices and pitfalls in inferring neural representations. Neuroimage 174:340–351.

Poulsen AT, Kamronn S, Dmochowski J, Parra LC, Hansen LK (2017) EEG in the classroom: Synchronised neural recordings during video presentation. Sci Rep 7:43916.

Power AJ, Foxe JJ, Forde EJ, Reilly RB, Lalor EC (2012) At what time is the cocktail party? A late locus of selective attention to natural speech. Eur J Neurosci 35:1497–1503.

Prendergast G, Johnson SR, Green GGR (2010) Temporal dynamics of sinusoidal and non-sinusoidal amplitude modulation. Eur J Neurosci 32:1599–1607.

Rabbitt PMA (1968) Channel-capacity, intelligibility and immediate memory. Q J Exp Psychol 20:241–248.

Rabinowitz NC, Willmore BDB, Schnupp JWH, King AJ (2011) Contrast Gain Control in Auditory Cortex. Neuron 70:1178–1191.

Rajagovindan R, Ding M (2010) From Prestimulus Alpha Oscillation to Visual-evoked Response: An Inverted-U Function and Its Attentional Modulation. J Cogn Neurosci 23:1379–1394.

Richard JP, Leppelsack H-J, Hausberger M (1995) A rapid correlation method for the analysis of spectro-temporal receptive fields of auditory neurons. J Neurosci Methods 61:99–103.

Rimmele J, Jolsvai H, Sussman E (2011) Auditory Target Detection Is Affected by Implicit Temporal and Spatial Expectations. J Cogn Neurosci 23:1136–1147.

Rockstroh B, Elbert T, Birbaumer N, Lutzenberger W (1982) Brain Potentials and Behavior. Urban & Schwarzenberg.

Rönnberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, Dahlström Ö, Signoret C, Stenfelt S, Pichora-Fuller KM, Rudner M (2013) The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front Syst Neurosci 7:31.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc London B Biol Sci 336:367–373.

Ru P (2001) Multiscale Multirate Spectro-Temporal Auditory Model. Unpublished doctoral dissertation, University of Maryland College Park.

Rudner M, Rönnberg J, Lunner T (2011) Working Memory Supports listening in Noise for Persons with Hearing Impairment. J Am Acad Audiol 22:156–167.

Salmelin R, Hari R (1994) Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. Neuroscience 60:537–550.

Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. PLOS Comput Biol 10:e1003412.

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci 22:2917–2926.

Sawaki R, Luck SJ (2010) Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. Attention, Perception, Psychophys 72:1455–1470.

Schubert ED, Schultz MC (1962) Some Aspects of Binaural Signal Selection. J Acoust Soc Am 34:844–849.

Shamma S, Elhilali M, Ma L, Micheyl C, Oxenham AJ, Pressnitzer D, Yin P, Xu Y (2013) Temporal coherence and the streaming of complex sounds. Adv Exp Med Biol 787:535–543.

Shannon CE (1948) A Mathematical theory of Communication. Bell Syst Tech J 27:379–423.

Shannon R V, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech Recognition with Primarily Temporal Cues. Science 270:303–304.

Shapiro I (1979) Evaluation of Relationship Between Hearing Threshold and Loudness Discomfort Level in Sensorineural Hearing Loss. J Speech Hear Disord 44:31–36.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182–186.

Smeds K, Wolters F, Rung M (2015) Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. J Am Acad Audiol 196:183–196.

Smulders FTY, Oever S, Donkers FCL, Quaedflieg CWEM, Ven V (2018) Single-trial log transformation is optimal in frequency analysis of resting EEG alpha. Eur J Neurosci 48:2585–2598.

Snyder JS, Gregg MK, Weintraub DM, Alain C (2012) Attention, awareness, and the perception of auditory scenes. Front Psychol 3:15.

Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive Top-Down Integration of Prior Knowledge during Speech Perception. J Neurosci 32:8443–8453.

Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M (2017) Is predictability salient? A study of attentional capture by auditory patterns. Philos Trans R Soc London B Biol Sci 372:20160105.

Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. Science 240:338–340.

Strauß A, Wöstmann M, Obleser J (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. Front Hum Neurosci 8:350.

Sweet RA, Dorph-Petersen K-A, Lewis DA (2005) Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J Comp Neurol 491:270–289.

Tallon-Baudry C, Kreiter AG, Bertrand O (1999) Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. Vis Neurosci 16:449–459.

Tavabi K, Obleser J, Dobel C, Pantev C (2007) Auditory evoked fields differentially encode speech features: an MEG investigation of the P50m and N100m time courses during syllable processing. Eur J Neurosci 25:3155–3162.

Teki S, Chait M, Kumar S, Shamma S, Griffiths TD (2013) Segregation of complex acoustic scenes based on temporal coherence. eLife 2:e00699.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.

Thwaites A, Glasberg BR, Nimmo-Smith I, Marslen-Wilson WD, Moore BCJ (2016) Representation of Instantaneous and Short-Term Loudness in the Human Cortex. Front Neurosci 10:1–11.

Tiitinen HT, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Näätänen R (1993) Selective attention enhances the auditory 40-Hz transient response in humans. Nature 364:59–60.

Treisman AM (1960) Contextual cues in selective listening. Q J Exp Psychol 12:242–248.

Tsuchida T, Cottrell GW (2012) Auditory Saliency Using Natural Statistics. Proc Annu Meet Cogn Sci Soc 34:1048–1053.

Vachon F, Labonté K, Marsh JE (2017) Attentional capture by deviant sounds: A noncontingent form of auditory distraction? J Exp Psychol Learn Mem Cogn 43:622–634.

Van Drongelen W, Yuchtman M, Van Veen BD, Huffelen AC Van (1994) A Spatial Filtering Technique to Detect and Localize Multiple Sources in the Brain. Brain Topogr 9:39–49.

Van Eyndhoven S, Francart T, Bertrand A (2016) EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. IEEE Trans Biomed Eng:https://doi.org/10.1109/TBME.2016.2587382.

Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of Brain Electrical Activity via Linearly Constrained Minimum Variance Spatial Filtering. IEEE Trans Biomed Eng 44:867–880.

Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T (2018) Speech intelligibility predicted from neural entrainment of the speech envelope. J Assoc Res Otolaryngol 19:181–191.

Venezia JH, Thurman SM, Richards VM, Hickok G (2019) Hierarchy of speech-driven spectrotemporal receptive fields in human auditory cortex. Neuroimage 186:647–666.

Verschueren E, Somers B, Francart T (2018) Neural envelope tracking as a measure of speech understanding in cochlear implant users. Hear Res:https://doi.org/10.1101/469643.

Vissers ME, van Driel J, Slagter HA (2016) Proactive, but Not Reactive, Distractor Filtering Relies on Local Modulation of Alpha Oscillatory Activity. J Cogn Neurosci 28:1964–1979.

Waldrop MM (2016) The chips are down for Moore's law. Nat News 530:144.

Wang D, Brown GJ (2006) Computational Auditory Scene Analysis: Principles, Algorithms, and Applications. Wiley-IEEE Press.

Wang X, Lu T, Liang L (2003) Cortical processing of temporal modulations. Speech Commun 41:107–121.

Wang Y, Zhang J, Ding N, Zou J, Luo H (2018) Prior Knowledge Guides Speech Segregation in Human Auditory Cortex. Cereb Cortex bhy052:1–11.

Weisz N, Hartmann T, Müller N, Obleser J (2011) Alpha Rhythms in Audition: Cognitive and Clinical Perspectives. Front Psychol 2:73.

Widmann A, Schröger E, Maess B (2014) Digital filter design for electrophysiological data – a practical approach. J Neurosci Methods 250:34–46.

Willmore BDB, Cooke JE, King AJ (2014) Hearing in noisy environments: noise invariance and contrast gain control. J Physiol 16:3371–3381.

Wingfield A, Tun PA, McCoy SL (2005) Hearing Loss in Older Adulthood: What It Is and How It Interacts with Cognitive Performance. Curr Dir Psychol Sci 14:144–148.

Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci U S A 90:8722–8726.

Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5:1–7.

Wong DDE, Fuglsang SA, Hjortkjær J, Ceolini E, Slaney M, de Cheveigné A (2018) A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. Front Neurosci 12:531.

Woolgar A, Jackson J, Duncan J (2016) Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. J Cogn Neurosci 28:1433–1454.

Worden MS, Foxe JJ, Wang N, Simpson G V (2000) Anticipatory Biasing of Visuospatial Attention Indexed by Retinotopically Specific α-Bank Electroencephalography Increases over Occipital Cortex. J Neurosci 20:RC63.

Wöstmann M, Fiedler L, Obleser J (2017a) Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. Lang Cogn Neurosci 32.

Wöstmann M, Lim S, Obleser J (2017b) The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. Cereb Cortex 27:3307–3317.

Wöstmann M, Herrmann B, Maess B, Obleser J (2016) Spatiotemporal dynamics of auditory attention synchronize with speech. Proc Natl Acad Sci U S A 113:201523357.

Wöstmann M, Herrmann B, Wilsch A, Obleser J (2015) Neural Alpha Dynamics in Younger and Older Listeners Reflect Acoustic Challenges and Predictive Benefits. J Neurosci 35:1458–1467.

Wöstmann M, Schmitt L-M, Obleser J (2018) Does closing the eyes enhance auditory attention? Eye closure increases attentional alpha-power modulation but not listening performance. bioRxiv:https://doi.org/10.1101/455675.

Zatorre RJ, Belin P, Penhune V (2002) Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37-46. Trends Cogn Sci 6:37–46.

Zink R, Proesmans S, Bertrand A, Van Huffel S, De Vos M (2017) Online detection of auditory attention with mobile EEG: closing the loop with neurofeedback. bioRxiv:https://doi.org/10.1101/218727.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991.

Zobel BH, Freyman RL, Sanders LD (2015) Attention is critical for spatial auditory object formation. Attention, Perception, Psychophys 77:1998–2010.

Zschocke S (2012) Klinische Elektroenzephalographie (Hansen H-C, ed)., 3rd ed. Berlin Heidelberg: Springer-Verlag GmbH.

List of figures

6 Summary

6.1 Introduction

In natural environments, multiple objects compete for our attention. Since cognitive resources are limited, the incoming information must be reduced. In theory, this reduction is achieved by attentional filtering. Our ability to selectively attend one sound source and ignore others (i.e., *the cocktail-party problem*; Cherry, 1953) has been a main body of research in the last decades. Psychophysical studies showed which features of a sound lead to the formation of auditory objects (i.e., *auditory scene analysis*; Bregman, 1990) and that selective attention can be allocated to different stimulus representations along the hierarchy of the auditory pathway (*early vs. late selection*; Broadbent, 1958; Treisman, 1960). The fact that even initially unattended stimuli can capture our attention led to the dichotomous concept of *bottom-up* attention (i.e., stimulus-driven) and *top-down* attention (i.e., goal-driven; Egeth and Yantis, 1997). However, how selective filtering is neurally established is still under investigation.

The neural selective processing of continuous speech recently moved into the focus of research. This research profits from the high temporal resolution of electrophysiological methods such as electroencephalography (EEG) and magnetoencephalography (MEG). Two neural signatures of auditory attention were investigated in detail: First, the *neural tracking* of speech refers to different methods that quantify the neural phase-locking to the (spectro-) temporal fluctuations of speech (e.g., Ding and Simon, 2012). Second, the power of induced neural oscillations around 10 Hz (i.e., *alpha power;* Berger, 1932) was proposed to be indicative of the distribution of neural resources through inhibition of brain regions or neural pathways processing irrelevant information (Klimesch et al., 2007; Jensen and Mazaheri, 2010). It was shown that out of a mixture of two talkers, a clean spectro-temporal representation of the attended talker is established in auditory cortex. It was also shown that the modulation of *alpha power* indicates the spatial focus of attention (Kerlin et al., 2010) and the demand for attentional control (Obleser and Weisz, 2012; Wöstmann et al., 2017b). However, a comprehensive understanding of the functional link between *neural tracking* of speech and the modulation of *alpha power* has not been established yet. Here we investigated the simultaneous attentional modulation of *neural tracking* and *alpha power*.

The finding that the attentional focus of a listener can be estimated from the EEG within a minute or so fueled the development of neurally steered hearing aids (e.g., O'Sullivan et al., 2014). Consequently, this approach was investigated based on a reduced set of hearing-aid-compatible EEG electrodes, such as electrodes placed around the ear (e.g., Mirkovic et al., 2016). While above-chance detection based on *neural tracking* could be achieved, other neural measures such as the modulation of *alpha power* might also reveal information about a listener's attentional focus, valuable for neural steering of a hearing aid. Here we investigated the neural signatures of auditory attention captured by electrodes placed inside the ear canal (in-ear EEG).

The goal of this thesis was to investigate the signatures of auditory attention under continuously varying listening conditions. At the core of this thesis, two concurrent talkers were presented while subjects were asked to attend one of them (signal) and to ignore the other (noise). We continuously varied the signal-to-noise (SNR) and the location of the talkers to manipulate the demand on attentional control. We applied and refined *forward encoding models*, which allowed us to predict the neural response to continuous speech as well as to detect the attentional focus of a listener. Based on the predictions, we traced the cortical representation (i.e., *neural tracking*) of the attended and the ignored talker, respectively. We disentangled the impact of *bottom-up*-driven versus *top-down*-attentional modulation (i.e., *neural selectivity*). We hypothesized that both, the *neural tracking* and *alpha power* would be modulated by the manipulation of the listening condition. First, we expected that the differential *neural tracking* between the attended and the ignored talker shows early *bottom-up*, but late *top-down* modulation. Second, we expected to find increased *alpha power* during more adverse conditions, indicating *top-down* suppression of task-irrelevant brain regions or, under spatial segregation, suppression of irrelevant neural pathways.

This thesis shows that the *neural tracking* of speech is a more prominent neural signature of auditory attention compared to *alpha power*. In the first part, we show that the *neural tracking* is primarily shaped by *top-down* attention, resulting in suppressed responses to the ignored talker. Under most adverse conditions, late *neural tracking* of the ignored talker indicates its *top-down*-controlled suppression. The modulation of *alpha power* was not following the expected pattern. Neither the SNR nor the location of the talkers predicted the modulation of *alpha power*. In the second part of this thesis, we show that phase-locked neural responses to auditory stimuli can be

recorded from in-ear EEG. We show that a listener's focus of attention can be detected with in-ear EEG based on the *neural tracking* of speech. We replicate this finding and additionally show that the increased late tracking of the ignored talker is also indicated by in-ear EEG.

6.2 Experiments and results

In studies 1–3, we investigated how the neural signatures of auditory attention to continuous speech are shaped by *bottom-up* and *top-down* attention. We focused our analysis on the *neural tracking* of speech and the modulation of *alpha power*. The adversity of the listening condition was manipulated by the variation of the signal-to-noise ratio (SNR) as well as the location of the talkers.

In study 1, we stochastically varied the SNR between a to-be-attended and a to-be-ignored talker, while we recorded the EEG of N=18 normal hearing subjects. We investigated the attention- and SNR-dependent *neural tracking* of speech. We show that the *neural tracking* of speech is highly controlled by *top-down* attention, resulting in suppressed neural response to the ignored talker in temporal cortical regions. Importantly, in the most adverse listening condition, the late neural selective processing of the ignored talker plays a crucial role for the overall selective processing of the *auditory scene*, brought up by *top-down* suppression of the ignored talker. This late neural processing of the ignored talker is accomplished by fronto-parietal brain regions, which highlights that it is brought up by *top-down* attentional control at a late stage.

Study 2 was based on the same experimental data as study 1. In contradiction to our hypothesis, we show that *alpha power* does not indicate the current demand for *top-down* attentional control. Further exploratory analysis did not reveal a clear relationship between *alpha power* modulation and the SNR. A direct comparison of *neural tracking* versus *alpha power* in terms of their modulation by attention showed that during concurrent, continuous speech, neural signatures of auditory attention are predominantly emerging from the phase-locked neural responses (i.e., *neural tracking*).

In study 3, we extended the experimental design of study 1. In addition, we stochastically varied the location of the talkers, such that we expected that the lateralization of *alpha power* should indicate the attended location when talkers are situated at different positions. We recorded the EEG of N=25 subjects. In contradiction to our hypothesis, neither the location of the talkers,

the SNR, nor the interaction of the two were found to modulate *alpha power* consistently. Particularly, we showed that neither the modulation of *alpha power* at single EEG electrode positions nor the whole-scalp hemispherical imbalance (i.e., *alpha power* lateralization) are indicative of the location of the attended talker.

In studies 4–6, we investigated whether the feasibility of a reduced set of EEG electrodes including in-ear EEG detects a listener's focus of auditory attention, such that in-ear EEG might be used to inform a neurally steered hearing aid.

In study 4, we presented sounds rich of spectro-temporal modulation, while in-ear EEG of N=6 subjects was recorded. We show that a single-subject's spectrally resolved neural responses can be extracted from two EEG electrodes placed around and inside the ear canal. However, we also show that an increased spectral resolution does not necessarily lead to a more precise prediction of the EEG signal.

In study 5, we presented concurrent dichotic tone streams and diotic speech to N=8 subjects, while scalp and in-ear EEG was recorded. We show that single-channel in-ear EEG electrode configurations capture the *neural tracking* of the auditory streams. The attentional focus could be detected within single subjects. This study suggests the feasibility of single-channel EEG configurations to be attached to a hearing as a basis for the neural steering.

In study 6, we used the experimental design of study 1, while the in-ear and scalp EEG of N=6 subjects was recorded. We replicated results of study 5 by showing that single-channel electrode configurations detect a listener's focus of attention. Furthermore, under the most adverse condition, we found a similar increase of late *neural selectivity* in the response to the ignored talker, as found previously in study 1.

6.3 Discussion

Our findings demonstrate that the *neural tracking* of speech is shaped by the *top-down* attentional goal of a listener and that the morphology of the neural responses reveal distinct signatures of the demand for attentional control. We also showed that the attention-dependent *neural tracking* is a potential source of information to steer a hearing aid. We could not draw a clear picture on the attentional modulation of *alpha power*. Based on our findings, the

involvement of *alpha power* in *top-down* attentional control stays inconclusive, but we can conclude that *neural tracking* of speech is much more prominent neural signature of auditory attention in EEG.

One recurrent question within the discussion of our results is why *alpha power* modulation was observed in some (e.g., Wöstmann et al., 2016), but not our studies. We argue here that the *top-down neural tracking* of speech and the modulation of *alpha power* are two distinct neural strategies rather than two sides of the same coin. In what follows, we will discuss the filter strategies and why their involvement might strongly depend on the experimental design.

How is the attention-dependent *neural tracking* achieved in temporal brain regions? Most likely, the clean representation of the attended talker is brought up by the tuning of neurons to the spectro-temporal characteristics of the attended talker (Fritz et al., 2007; Lakatos et al., 2013). In our studies, we always presented concurrent talkers of different gender, such that their differential spectro-temporal characteristics provided a distinct feature for such a filter located in auditory cortex (Mesgarani and Chang, 2012). Under most adverse conditions, we observed a signature of a late, most likely reactive *top-down* suppression of the ignored talker, a role we hypothetically ascribed to *alpha power* in advance. We argue that our task design did not challenge subjects in the way that dedicated *alpha power* neural strategies came into play, as can be found in trial-based designs. We conclude that attentional filtering was primarily achieved by spectro-temporal, proactive filtering in auditory cortex. Thus, if a spectro-temporal distinction between the talkers is not present (e.g., same talker on left and right ear; see Wöstmann et al., 2016), *alpha power* neural strategies might come into play.

We found that single channel in-ear EEG captures the auditory attentional focus based on the attention-dependent *neural tracking* of the talkers. This was even achieved under most adverse listening conditions. We argue that real-life listening scenarios bear further challenges that cannot be resolved in the laboratory (Hamilton and Huth, 2018), such that prototypes of neurally steered hearing should be tested in more realistic environments. This might reveal additional neural signatures of auditory attention and contribute to a better understanding of the functional role of *alpha power* modulation.

In sum, the *neural tracking* of speech was found to be the most prominent signature of *top-down* auditory attention, which adapts to the current listening conditions. The role of *alpha power* modulation as a signature of *top-down* attention to continuous speech was not confirmed by this thesis. It is up to further studies to close the gap between trial-based designs gathering enough behavioral data and continuous, ecologically valid designs allowing natural behavior.

7 Zusammenfassung

7.1 Einführung

In natürlichen Umgebungen konkurrieren viele Reize gleichzeitig um unsere Aufmerksamkeit. Da kognitive Ressourcen begrenzt sind, müssen die eingehenden Informationen reduziert werden. Theoretisch wird diese Reduktion durch den Filter der Aufmerksamkeit erreicht. Unsere Fähigkeit, eine Schallquelle selektiv zu beachten und andere zu ignorieren (*Cocktailpartyproblem*; Cherry, 1953), war ein Hauptgegenstand der Forschung der letzten Jahrzehnte. Psychophysikalische Studien zeigten, welche Merkmale eines Klanges zur Bildung auditorischer Objekte führen (*Auditory scene analysis*; Bregman, 1990), und dass selektive Aufmerksamkeit an verschiedenen Repräsentationen des Reizes entlang der Hierarchie der Hörbahn zum Tragen kommt (frühe vs. späte Selektion; Broadbent, 1958; Treisman, 1960). Auch zunächst unbeachtete Reize können unsere Aufmerksamkeit erregen, was zu dem dichotomen Konzept von *Bottom-Up*-Aufmerksamkeit (durch den reiz getrieben) und *Top-Down*-Aufmerksamkeit (durch Ziele getrieben; Egeth und Yantis, 1997) geführt hat. Wie die selektive Filterung neural stattfindet, ist jedoch weiterhin Gegenstand der Forschung.

Die neurale, selektive Verarbeitung von kontinuierlich gesprochener Sprache ist in den letzten Jahren zunehmend Gegenstand der Forschung geworden. Dies wurde begünstigt durch die hohe zeitliche Auflösung elektrophysiologischer Verfahren wie Elektroenzephalographie (EEG) und Magnetoenzephalographie (MEG). Zwei neurale Merkmale auditiver Aufmerksamkeit waren dabei unter genauerer Beobachtung: Einerseits zeichnen sich die Muster der Zeit- und Frequenzmodulation von Sprache in Form einer phasentreuen neuralen Antwort im M-/EEG ab (d.h. *neurales Tracking*; z. B. Ding und Simon, 2012). Zum anderen weisen induzierte neurale Wellen um 10 Hz (d. H. *Alpha-Wellen*; Berger, 1932) auf die Verteilung neuraler Ressourcen durch Unterdrückung von Gehirnarealen oder neuraler Bahnen hin (Klimesch et al., 2007; Jensen und Mazaheri, 2010). Es wurde gezeigt, dass aus einer Mischung von zwei Sprechern eine bereinigte Zeit-Frequenz-Repräsentation des beachteten Sprechers im auditorischen Kortex herausgearbeitet wird. Es wurde auch gezeigt, dass die Amplitudenmodulation von *Alpha-Wellen* auf den räumlichen Fokus der Aufmerksamkeit (Kerlin et al., 2010) und die Steuerung von Aufmerksamkeit hinweist (Obleser und Weisz, 2012; Wöstmann et al., 2017b). Ein umfassendes

Verständnis der funktionalen Verbindung zwischen *neuralem Tracking* und der Modulation von *Alpha-Wellen* ist jedoch noch nicht gegeben. In der vorliegenden Arbeit wurde die gleichzeitige Modulation von *neuralem Tracking* und *Alpha-Wellen* durch Aufmerksamkeit untersucht.

Der Befund, dass der Fokus auditiver Aufmerksamkeit aus dem EEG innerhalb von ca. einer Minute geschätzt werden kann, hat die Entwicklung neural gesteuerter Hörgeräte angeregt (z. B. O'Sullivan et al., 2014). Folglich wurde diese Methode mit einer reduzierten Anordnung von EEG-Elektroden umgesetzt, die z.B. um das Ohr platziert werden (z. B. Mirkovic et al., 2016). Während die Erkennung von Aufmerksamkeits anhand des *neuralen Trackings* durchgeführt wurde, könnten andere neurale Merkmale wie die Modulation der *Alpha-Wellen* weitere Informationen über den Aufmerksamkeitszustand eines Hörers liefern, welche zur neuralen Steuerung von Hörgeräten beitragen könnten. In dieser Arbeit wurden die neuralen Merkmale auditiver Aufmerksamkeit mit im Gehörgang platzierten Elektroden untersucht (in-ear-EEG).

Das Ziel dieser Arbeit war es, die Merkmale auditiver Aufmerksamkeit unter veränderlichen Hörbedingungen zu untersuchen. Bei den Kernstudien dieser Arbeit wurden zwei Sprecher gleichzeitig präsentiert, während die Probanden instruiert wurden, einen der beiden Sprecher zu beachten (Signal) und den anderen oder die andere zu ignorieren (Rauschen). Wir haben das Signal-Rausch-Verhältnis (signal-to-noise ratio; SNR) und die Position der Sprecher kontinuierlich variiert, um den Bedarf nach Aufmerksamkeitssteuerung zu beeinflussen. Wir Enkodierungsmodelle angewendet, mit denen wir die neurale Antwort auf kontinuierliche Sprache vorhersagen und den Fokus der Aufmerksamkeit eines Zuhörers detektieren konnten. Basierend auf den Vorhersagen verfolgten wir die kortikale Repräsentation (d. h. *Neurales Tracking*) des beachteten bzw. des ignorierten Sprechers. Wir haben die anteiligen *Bottom-Up*-getriebenen und *Top-Down*-gesteuerten (d.h. *Neurale Selektivität*) Einflüsse auf die neurale Antwort analysiert. Wir erwarteten, dass sowohl das *neurale Tracking* als auch Amplitude der *Alpha-Wellen* durch diese Manipulation der Hörbedingungen beeinflusst werden. Zunächst erwarteten wir, dass das differentielle *neurale Tracking* zwischen dem beachteten und dem ignorierten Sprecher eine frühe *Bottom-Up*-, jedoch eine späte *Top-Down*-Modulation zeigt. Zweitens haben wir erwartet, dass die Amplitude der *Alpha-Wellen* unter schwierigeren Bedingungen größer ausfällt, was eine Unterdrückung der zu ignorierenden Sprechers zur Verhinderung von *Bottom-Up*-getriebener Aufmerksamkeit bedeuten würde. Des Weiteren

erwarteten wir einen Zusammenhang zwischen *neuralem Tracking* und der Modulation der *Alpha-Wellen* zu finden.

Diese Arbeit zeigt, dass das *neurale Tracking* von Sprache ein valideres neurales Merkmal auditiver Aufmerksamkeit im Vergleich zu *Alpha-Wellen* darstellt. Im ersten Teil zeigen wir, dass das *neurale Tracking* in erster Linie von der *Top-Down*-Aufmerksamkeit geprägt wird, was zur Unterdrückung der neuralen Antwort auf den ignorierten Sprecher führt. Unter den schwierigsten akustischen Bedingungen weist ein spätes *neurales Tracking* des ignorierten Sprechers auf seine *Top-down*-gesteuerte Unterdrückung hin. Die Modulation der *Alpha-Wellen* folgte nicht dem erwarteten Muster. Weder das SNR noch die Position der Sprecher ließen eine Vorhersage der zu erwartenden *Alpha-Wellen*-Amplitude zu. Im zweiten Teil dieser Arbeit wird zum einen gezeigt, dass die phasentreue neurale Antwort auf Hörreize vom EEG im Ohr aufgenommen werden können. Es wird des Weiteren gezeigt, dass der Fokus eines Zuhörers mit in-ear-EEG auf der Grundlage des *neuralen Trackings* von Sprache ermittelt werden kann. Wir replizieren diesen Befund und zeigen zusätzlich, dass das späte neurale Tracking des ignorierten Sprechers auch mit In-Ear-EEG aufgezeichnet werden kann.

7.2 Experimente und Ergebnisse

In den Studien 1–3 wurde untersucht, wie die neuralen Merkmale der auditiven Aufmerksamkeit auf kontinuierlicher Sprache durch *Bottom-Up*- und *Top-Down*-Aufmerksamkeit beeinflusst werden. Wir konzentrierten unsere Analysen auf das *neurale Tracking* von Sprache und die Modulation von *Alpha-Wellen*. Die Hörbedingung wurde durch die Veränderung des SNRs sowie durch den Ort der Sprecher beeinflusst.

In Studie 1 haben das SNR zwischen einem zu beachtendem und einem zu ignorierenden Sprecher stochastisch variiert, während wir das EEG von N=18 normalhörenden Probanden aufzeichnet haben. Wir untersuchten das aufmerksamkeits- und SNR-abhängige *neurale Tracking* von Sprache. Diese Studie zeigt, dass das *neurale Tracking* von Sprache in hohem Maße von der *Top-Down*-Aufmerksamkeit beeinflusst wird, was zu einer unterdrückten neuralen Antwort auf den ignorierten Sprecher in temporalen Gehirnregionen führt. Wichtig ist, dass in schlechtesten Hörbedingungen die späte neurale selektive Verarbeitung des ignorierten Sprechers eine entscheidende Rolle für die gesamte selektive Verarbeitung der Hörsituation spielt, was auf

eine *Top-Down*-Unterdrückung des ignorierten Sprechers zurückzuführen ist. Diese späte neurale Verarbeitung des ignorierten Sprechers findet in fronto-parietalen Hirnregionen statt, was unterstreicht, dass sie durch eine späte *Top-down*-gesteuerte Aufmerksamkeitssteuerung hervorgebracht wird.

Studie 2 basiert auf dem Datensatz von Studie 1. Entgegen unserer Erwartung zeigen wir, dass *Alpha-Wellen* nicht die Unterdrückung des zu ignorierenden Sprechers abbilden und damit kein Merkmal von Aufmerksamkeitssteuerung darstellen. Weitere explorative Analysen ergaben keinen eindeutigen Zusammenhang zwischen der Modulation von *Alpha-Wellen* und dem SNR. Ein direkter Vergleich der im *neuralem Tracking* beobachteten Selektivität und der Modulation der *Alpha-Wellen* zeigte, dass während der kontinuierlichen Sprache neurale Merkmale auditiver Aufmerksamkeit vorwiegend aus dem *neuralen Tracking* hervorgehen.

In Studie 3 wurde das experimentelle Design von Studie 1 um die stochastische Bewegung der erweitert. Wir erwarteten eine Lateralisierung der *Alpha-Wellen*, welche die Position des beachteten Sprechers anzeigt. Wir haben das EEG von N=25 Probanden aufgenommen. Entgegen unserer Hypothese konnten wir weder die Position der Sprecher, das SNR, noch eine Interaktion feststellen. Insbesondere haben wir gezeigt, dass weder die Modulation der *Alpha-Wellen* an einzelnen EEG-Elektroden noch die *Alpha-Wellen*-Lateralisierung über den gesamten Kopf die Position des beachteten Sprechers anzeigen.

In den Studien 4–6 haben wir untersucht, ob eine reduzierte Konfiguration von EEG-Elektroden, einschließlich in-ear-EEG, den Fokus der Aufmerksamkeit des Hörers anzeigt. Folglich könnte in-ear-EEG zur neuralen Steuerung eines Hörgeräts verwendet werden.

In Studie 4 wurden Geräusche reich an Zeit-Frequenz-Modulationen präsentiert und dabei das in-ear-EEG von N=6 Probanden aufgezeichnet. Wir zeigen, dass die spektral aufgelöste neurale Antwort im einzelnen Probanden aus zwei im und um das Ohr platzierten EEG-Elektroden extrahiert werden kann. Wir zeigen jedoch auch, dass eine erhöhte spektrale Auflösung nicht unbedingt zu einer genaueren Vorhersage des EEG-Signals führt.

In Studie 5 wurden N=8 Probanden gleichzeitige dichotische Tonfolgen und diotische Sprache präsentiert, während Skalp- und in-ear-EEG aufgezeichnet wurden. Wir zeigen, dass einkanalige EEG-Elektrodenkonfigurationen das *neurale Tracking* der auditiven Reize erfassen. Der

Aufmerksamkeitsfokus konnte im einzelnen Probanden ermittelt werden. Diese Studie zeigt, dass einkanalige EEG-Konfigurationen als Grundlage für die neurale Steuerung von Hörgeräten verwendet werden können.

In Studie 6 wurde das experimentelle Design von Studie 1 verwendet, während Skalp- und in-ear-EEG von N=6 Probanden aufgezeichnet wurde. Wir konnten zum einen Studie 5 replizieren, indem wir zeigen, dass einkanalige Elektrodenkonfigurationen den Aufmerksamkeitsfokus des Hörers anzeigen. Zum anderen zeigen wir, dass unter den schwierigsten Hörbedingungen eine Zunahme des späten *neuralen Trackings* des ignorierten Sprechers vorliegt, wie wir sie zuvor auch in Studie 1 beobachtet haben.

7.3 Diskussion

Unsere Ergebnisse zeigen, dass das *neurale Tracking* von Sprache durch den *Top-Down*-Fokus eines Zuhörers geformt wird und dass der Verlauf der neuralen Antwort unterschiedliche Merkmale der Aufmerksamkeitssteuerung aufzeigt. Wir haben auch gezeigt, dass das aufmerksamkeitsabhängige *neurale Tracking* ein zuverlässiges Merkmal für die Steuerung eines Hörgeräts ist. Unsere Ergebnisse lassen keinen klaren Schluss auf die aufmerksamkeitsabhängige Modulation von *Alpha-Wellen* zu. Basierend auf unseren Ergebnissen ist die Bedeutung von Alpha-Power für die *Top-Down*-Aufmerksamkeitssteuerung nicht eindeutig, aber wir können daraus schließen, dass das *neurale Tracking* von Sprache ein stärkeres neurales Merkmal der auditiven Aufmerksamkeit im EEG ist.

Eine wiederkehrende Frage in der Diskussion unserer Ergebnisse ist, warum die Modulation von *Alpha-Wellen* in einigen, aber nicht in anderen Studien beobachtet wurde. Wir argumentieren hier, dass das *neurale Tracking* von Sprache und die Modulation der *Alpha-Wellen* zwei verschiedene neurale Strategien und nicht zwei Merkmale der gleichen Strategie darstellen. Folglich diskutieren wir die potenzielle Filterstrategien und warum ihre neurale Anwendung stark vom experimentellen Design abhängen kann.

Wie wird das aufmerksamkeitsabhängige *neurale Tracking* in temporalen Hirnregionen ermöglicht? Wahrscheinlich ist, dass eine bereinigte Repräsentation des beachteten Sprechers durch die Abstimmung der Neuronen auf die zeitlich-spektralen Eigenschaften des beachteten Sprechers realisiert wird (Fritz et al., 2007; Lakatos et al., 2013). In unseren Studien haben wir stets

Sprecher unterschiedlichen Geschlechts gleichzeitig präsentiert, sodass ihre unterschiedlichen zeitlich-spektralen Eigenschaften ein eindeutiges Merkmal für die Einstellung eines solchen Filters im auditorischen Cortex sein könnte (Mesgarani und Chang, 2012). Unter den schwierigsten Hörbedingungen haben wir ein Merkmal einer späten, höchstwahrscheinlich reaktiven Unterdrückung des ignorierten Sprechers beobachtet, eine Funktion, die wir von *Alpha-Wellen* erwartet haben. Wir argumentieren, dass die absolvierung unserer Höraufgaben die Modulation von *Alpha-Wellen* nicht in dem Maß verlangten, wie dies bei Versuchen mit kurzen Durchgängen der Fall ist. Wir schließen daraus, dass die Aufmerksamkeitsfilterung in erster Linie durch zeitlich-spektrale, proaktive Filterung im auditorischen Kortex erreicht wurde.

Wir zeigen, dass anhand von in-ear-EEG der Fokus auditiver Aufmerksamkeit basierend auf dem *neuralen Tracking* der Sprecher ermittelt werden kann, was selbst unter schwierigsten Hörbedingungen erreicht wurde. Wir sind der Meinung, dass reale Hörszenarien weitere Herausforderungen mit sich bringen, die im Labor nicht simuliert werden können (Hamilton und Huth, 2018), sodass Prototypen von neural gesteuerten Hörgeräten in realistischeren Umgebungen getestet werden sollten. Dies könnte weitere neurale Merkmale auditorischer Aufmerksamkeit hervorbringen und zum besseren Verständnis der Funktion von *Alpha-Wellen* beitragen.

Zusammenfassend stellen wir fest, dass das *neurale Tracking* das dominantere Merkmal auditiver Aufmerksamkeit ist und eine neurale Anpassung an die aktuellen Hörbedingungen festzustellen ist. Die Funktion von *Alpha-Wellen* als Merkmal der Aufmerksamkeitssteuerung wurde durch diese Arbeit nicht bestätigt. Im Weiteren sollten Studien die Lücke zwischen durchgangsbasierten Designs, die genügend Verhaltensdaten hervorbringen, und kontinuierlichen, realistischeren Designs, die ein natürliches Verhalten abbilden, schließen.

8 Curriculum Vitae

Lorenz Fiedler
Dipl.-Wirtsch.-Ing. (FH) für Elektrotechnik
*25th of May, Jena, Germany

Employment

2016–18	Research assistant (PhD student), Department of Psychologie I, Research Group *Auditory Cognition,* Prof. Jonas Obleser, University of Lübeck, Germany. Project: "Towards the brain-informed, brain-controlled hearing aid"
2015	Research assistant, *Max-Planck Research Group Auditory Cognition*, *Max-Planck-Institute for Human Cognitive and Brain Sciences*, Leipzig, Germany Project: "Towards the brain-informed, brain-controlled hearing aid"
2014	Research assistant, *HTWK Leipzig*, *Research Group GeNuMedia*, Project: "Selective reception of acoustic media content"
2012	Research student: *Daimler AG*, Sindelfingen, Germany. Diploma thesis: "Synthetic sounds generation in cars with conventional engines"
2011	Intern, *Native Instruments GmbH,* Berlin, Gemany, Product Design/ Market Research/ Usability/ User Experience

Education

2007-2013	Diplom: Electrical engineering with management, *HTWK Leipzig*, Germany
1993-2006	Abitur (A-Level equivalent), Freie Waldorfschule Jena, Germany

Publications

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

Herbst SK, **Fiedler L**, & Obleser J, (2018). Tracking temporal hazard in the human electroencephalogram using a forward encoding model. *eNeuro* 5(2) 0017-18.2018

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

Wöstmann M, **Fiedler L**, & Obleser, J (2016). Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. *Lang. Cogn. Neurosci.* 135:52–6

Fiedler L, Obleser J, Lunner T, & Graversen C (2016) Ear-EEG allows extraction of neural responses in challenging listening scenarios - A future technology for hearing aids? In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. doi: 10.1109/EMBC.2016.7592020

Presentations

07/18	**Fiedler, L** (Hands-on workshop) Temporal response functions – Extraction of the neural response to continuous stimuli. *CuttingEEG*, Paris
05/18	**Fiedler L**, Wöstmann M, Herbst SK, & Obleser J (Invited talk) "Selective neural processing of speech under continuously varying listening conditions" *Psychologie und Gehirn (PuG),* Gießen, Germany
02/18	**Fiedler, L** (Invited talk) Signatures of auditory attention and listening effort in the human Electroencephalogram. *KU Leuven*, Leuven, Belgium, Host: Tom Francart
02/18	**Fiedler L**, Wöstmann M, Herbst SK, & Obleser J (Poster) Neural responses to concurrent speech reflect the emergence of an SNR-invariant representation of the attended talker. *Association for Research in Otolaryngology* (ARO), San Diego, US
4/17	**Fiedler, L** (Invited talk and Co-Chair) Auditive Aufmerksamkeit unter variierenden Hörbedingungen: neue methodische Ansätze, *DGKN*, Leipzig, Germany (Chair: Nathan Weisz)
2/17	**Fiedler L**, Wöstmann M, Herbst SK, & Obleser J (Invited talk) Scalp EEG predicts listeners' attentional focus and attentional demands under continuously varying signal-to-noise ratio, *DGA-Jahrestagung*, Ahlen, Germany (Chair: Stefan Debener)
11/16	**Fiedler L**, Wöstmann M, Herbst SK, Graversen C, Lunner T, & Obleser J (Poster) Scalp EEG predicts listeners' attentional focus and attentional demands under continuously varying signal-to-noise ratio. *Society for Neuroscience Conference* (SfN), San Diego, US
09/15	**Fiedler L**, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, & Obleser J (Poster) In-Ear-EEG indicates neural signatures of effortful auditory processing, *Society for Neuroscience Conference* (SfN), Chicago
10/15	**Fiedler L**, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, & Obleser J (Poster, *Best Poster Award*) In-Ear-EEG indicates neural signatures of effortful auditory processing *Entrainment of Brain Oscillations Conference*, Delmenhorst

Selbständigkeitserklärung

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde, und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind. Ich versichere, dass die vorliegende Arbeit in gleicher oder in ähnlicher Form keiner anderen wissenschaftlichen Einrichtung zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Es haben keine früheren erfolglosen Promotionsversuche stattgefunden. Die Promotionsordnung der *Universität zu Lübeck* ist mir bekannt und ich erkenne diese an.

Lübeck, den 21. Dezember 2018

_____

Lorenz Fiedler