Evaluating Listening Effort with Electroencephalography in Ecological Situations

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Tirdad Seifi Ala
14331011

Supervised by

Carina Graversen
Lauren V. Hadley
William M. Whitmer

Hearing Sciences – Scottish Section
Division of Clinical Neuroscience, School of Medicine
University of Nottingham

2021

Abstract

Listening effort is the deliberate allocation of mental resources to carry out a listening task. Spending effort can be rewarding if it yields desired results but to invest listening effort constantly will lead to negative consequences such as fatigue, difficulties in speech recall, and social disengagement, especially for hearing-impaired individuals.

Listening effort can be measured subjectively via questionnaire and/or objectively via behavioural and physiological measures. One of the most used measures for listening effort objectively is electroencephalography (EEG). EEG is a brain-imaging modality which provides excellent temporal resolution to study brain oscillations. EEG picks up electrical activities on the head (scalp EEG) or from the ears (ear-EEG) for investigating the brain in different cognitive tasks. Alpha power is one of the features that can be extracted from the EEG signal and has been widely used for measuring auditory and non-auditory effort. In this thesis, EEG will be recorded during effortful listening tasks, and alpha power will be extracted to investigate listening effort.

One of the downsides of the studies on listening effort is controlling (or not controlling) different parameters in the experimental environment which reduces generalisability to real-life scenarios. The aim of the current thesis is to measure listening effort in settings which are more ecologically valid compared to traditional laboratory scenarios. To increase ecological validity, motivation of individuals will be manipulated (through monetary reward) to account for personal factors, and different rooms will be simulated (through characterisation of reverberation time) to account for environmental factors in effortful tasks which involve listening to speech in noise. Single sentences or continuous discourse will be presented as stimuli speech to cover more realistic conversation in everyday life. Additionally, as a new and wearable technology, ear-EEG will be used in one study which has the potential to be used as an ambulatory EEG measurement in hearing aids.

The main hypothesis of the thesis is that alpha power increases with increased listening effort. The overall results of five different studies in ecologically valid situations showed that the pattern of alpha power can be opposite when listeners are presented with single-sentence stimuli or continuous discourse. In single-sentence paradigm, in line with the hypothesis, alpha power reflected listening effort. However, in continuous speech alpha power indicated performance of the individuals rather than expenditure of resources in the brain. These results suggest that applying a one-measure-for-all-scenarios approach when measuring listening effort is not reliable in a real-life setting.

Acknowledgements

I was 7 years old when my father took me to my first day of school. I was a scared boy back then - everything looked so unfamiliar to me. I felt that I was not cut out for this. But seeing my father standing and watching me from the sidelines gave me the power and heart warmth that I needed to take on the challenge.

22 years have passed since that day, and I am now writing to submit my doctorate thesis. Like on the first day of school, I have faced many unfamiliar situations during these 22 years. Each time, the love and support from my mother, father, and brother has given me the utmost strength to persevere. To provide the best learning environment for me, my family sacrificed in many ways. Probably the greatest sacrifice was that my mother encouraged me to pursue my PhD overseas, even though this meant we would live in different continents. I hope this acknowledgement letter can serve as a small gift for their sacrifices. This work, or anything that comes after, belongs to them, as I wouldn't be in this place in my life, not even close, without them.

I would also like to give huge thanks to my supervisors Bill, Lauren and Carina for their support and guidance. I have learnt many valuable lessons from them in these past 3 years. I would especially like to thank my former supervisor and my forever-mentor, Thomas, for all the things he did for me during my time at Eriksholm and after. Moving to a new country was a challenging experience for me in many aspects, but he was there for me like a father. I will always cherish the opportunity that I got to be his student and mentee.

I also met many wonderful and talented people at Eriksholm. Being there and learning from them day-in day-out was definitely the highlight of my young career. I would especially like to thank Dorothea and Gaby. The support I received from them both about work and on a personal level was incredible; I will be always indebted to them.

Last but not least, I would like to thank my dear friends, Beth and Hossein. They always listened to my rants about different issues in my PhD and gave me the best advice and insight. Having wonderful friends like them made the challenges I faced much easier to handle.

…. and to all the teachers, supervisors, friends, and colleagues who had my best interests in their heart. Even a single word or act of kindness of yours gave me extra motivation to be here. For that, I am grateful to you all.

Author's Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for any other academic qualification. If required, specific contributions are detailed at the beginning of each chapter. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765329. The funder had no role in the design of studies.

Tirdad Seifi Ala

November 2021

Table of Contents

List of Tables

List of Figures

Chapter 1
Concepts and Background

1.1 Introduction

In our everyday lives, we are constantly listening to the world around us. Listening is the key to effective communication with people around us. However, when the listening situation is difficult, it becomes an unpleasant experience that people would try to avoid it.

A reason for that is when listening situation is difficult, we need to invest more “listening effort”. Listening effort, similar to any other kind of mental effort, requires resources which should be drawn from the brain that could be otherwise used for other cognitive tasks. Spending effort can be rewarding if it yields desired results but to invest listening effort constantly will lead to many negative consequences as well. The most obvious one is that listening effort makes the person fatigued (McGarrigle et al., 2014). Spending too much effort also takes away “recall-speech” resources in the brain, which makes remembering things they hear more difficult (van Engen et al., 2012; Ward et al., 2016). Another downside is that when the listener has to spend effort constantly in order to hear sounds or speech, they prefer to disengage from social situations that require constant listening (e.g., Hallberg & Carlsson, 1993) .

Many listening situations in our daily lives (e.g., sitting in a noisy restaurant and listening to a conversation) can be demanding. Such situations make the listener exhausted due to fact that they need to invest listening effort to understand the speech. The effort, in this case, is to attend to the desired sound source(s) and suppress the unwanted interfering sound source(s) in the background. Such suppression requires complex processing in the brain which in return requires valuable cognitive resources that could be saved or used otherwise. With impaired auditory periphery more problems can occur as people with hearing loss have increased difficulty focusing on one sound source and ignoring background noise (Gatehouse & Akeroyd, 2006).

Traditionally, behavioural measures of listening difficulty have mainly focused on speech intelligibility in noise (e.g., Koelewijn et al., 2012; Nilsson et al., 1994; Plomp & Mimpen, 1979). However, the covert problem with listening effort is that even if speech intelligibility is optimal, other cognitive factors might be changing with the difficulty of the task. In fact, there are several studies that have shown objectively that listening effort can vary, even if speech intelligibility is unchanged (e.g., Houben et al., 2013; Sarampalis et al., 2009; see **Section 1.3** below). Therefore, studying listening effort beyond measures of speech intelligibility is necessary.

Measuring listening effort can be done subjectively and objectively. The main focus of this thesis will be on measuring listening effort objectively via electrical brain activities using electroencephalography (EEG). The advantage of using EEG is in its rich temporal information that can capture small electrical changes on the scalp within milliseconds order. In addition to that, EEG has the potential to be used as a wearable technology, possibly alongside hearing aids (see **Chapter 5**). This opens up new opportunities for automated hearing aids setting which can change according to the listening situation which is constantly shifting in our daily lives.

Therefore, it is also important to measure effort in listening situations which are relevant to real life. In order to do so, I will introduce different factors which are either personal (such as motivation) or environmental (such as conversation-like stimuli) that can be overlooked in studies related to listening effort. I will explore whether the listening effort measured with EEG data (specifically by alpha power) are affected by such personal and environmental factors, as well as task demand.

1.2 Listening effort

But what is exactly listening effort? A recent theoretical framework called Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016) incorporated Kahneman's Capacity Model of Attention (Kahneman, 1973), Brehm's Motivation Intensity Theory (MIT; Brehm & Self, 1989), and the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013) to listening studies. In FUEL, listening effort has been defined as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016, p. 10S). There is one important key phrase in this definition that needs more elaboration: *deliberate allocation of resources*.

Deliberate allocation of resources could be explained as one's willingness and motivation to invest effort in a task. Therefore, motivation is an integral part of FUEL model. The motivation aspect of this theory is mainly derived from MIT (Brehm &

Self, 1989). MIT postulated that people only put in more effort for cognitive tasks if their effort expenditure is perceived to yield appropriate benefit. In other words, effort investment is proportional to task demand, but the importance of performing the task also plays a role in how much the person is willing to invest effort. FUEL took this idea and expanded it by including cognition. Also, other confounding factors such as working memory capacity and fatigue can vary the measured effort during a listening task.

But what are the resources required for listening effort and why are they important? Resources are defined as means available for the execution of [listening] tasks in the brain (Wingfield, 2016). Their expenditure limits the available resources for other cognitive task such as recalling speech. The modulation of listening effort according to demand and motivation may be partly explained by increased allocation of working memory, which is particularly important for speech communication. In demanding listening situations, noisy representations of words lead to mismatches between the episodic memory and semantic memory in the working memory. Episodic memory is personal experiences tagged by time, place, and emotions. On the other hand, semantic memory is common knowledge such as meaning of the words and phonology. The mismatch between episodic and semantic memory calls for further explicit processing and storage capacity (Rönnberg et al., 2013). Thus these mismatches in speech representation require more cognitive processing resources for correct representation of words in working memory (Lemke & Besser, 2016; Peelle, 2018) which can lead to increased listening effort.

In our everyday lives, listening to a speech often includes listening to multiple sentences that may last for more than a few seconds. Therefore, it is a natural reaction that the person does not invest effort evenly over time. A person can adapt to, get fatigued by, or lose/gain motivation in the difficult situation during this long period. Therefore, it is more realistic to consider a time varying resource management of listening effort during an ongoing task (**Fig. 1.1**; Strauss & Francis, (2017)). For this reason, it is vital to study listening effort over an extended period for more ecological validity.

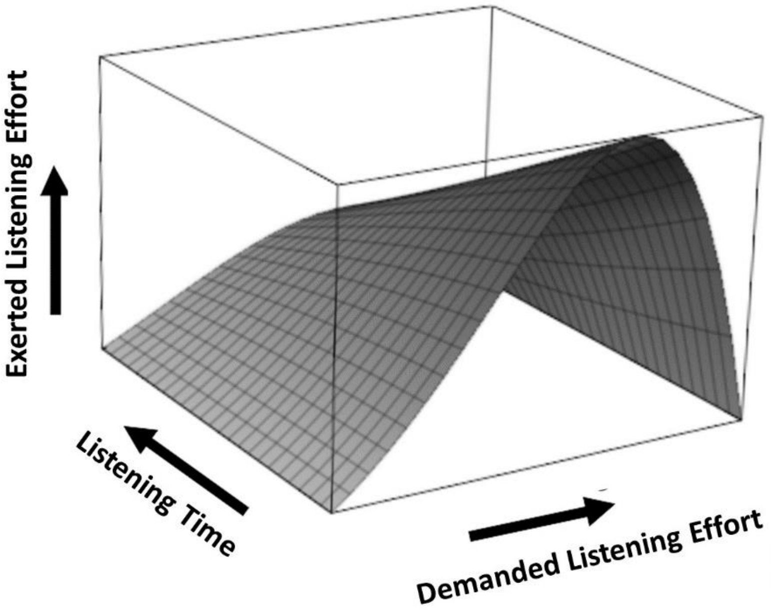


Fig. 1.1 Non-stationary changes of listening effort over an extended time period (modified figure from Strauss & Francis, (2017)).

1.3 Measuring listening effort

There are great number of ways to evaluate listening effort (see Alhanbali et al., 2019) which can be categorized into subjective and objective measures. Subjective measures are based on the perception of the person who performed the task and evaluated with questionnaire. On the other hand, objective measures are quantifiable outcomes that were collected from the person during performing the task. Objective measurements can themselves be divided into behavioural and physiological measures (see **Fig. 1.2**). Both subjective and objective (whether it is behavioural or physiological) measures have their own advantages and disadvantages. In this section, I briefly introduce each one and discuss their strengths and weaknesses when it comes to studying listening effort.

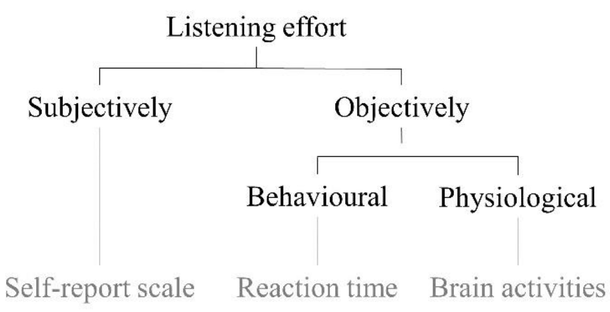


Fig. 1.2 Different methods of measuring listening effort with an example

1.3.1 Subjective measures

To evaluate listening effort subjectively, many researchers have developed different questionnaires, such as NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), Listening Effort Questionnaire-Cochlear Implant (LEQ-CI; Hughes et al., 2019), and Adaptive Categorical Listening Effort Scaling (ACALES; Krueger et al., 2017), that can be used for both normal hearing or hearing impaired. Subjective measures are based on listeners' own perception of the difficulty of the task. Collecting subjective measures are convenient and can be used in many various listening situations. There are few instances where subjective measures have been more sensitive to changes of effort compared to objective measures. For example, in a study by Johnson et al., (2015), participants were presented a set of words with four different signal-to-noise ratios (SNRs) either with high or low predictability context. The results showed that the self-reported effort (on a four-point scale) decreased linearly with increasing SNR for both predictability contexts compared to word recall which was used as a behavioural measure of listening effort. While this study showed the effectiveness of subjective measures, using word recall accuracy may have not been a reliable measure of listening effort. Such a measure can be highly dependent on memory span of the individuals, which may hinder studying the effects of listening effort (McGarrigle et al., 2014).

One of the reasons that subjective and objective measures do not align together is the fact that subjective effort is changed linearly with changing task demand, whereas objective measures often follow an inverted U-shaped pattern when a wide range of task demand is used. As an example, in the study by Zekveld and Kramer (2014), participants listened to speech-in-noise in 9 different SNRs from -36 dB SNR to -4 dB SNR in 4 dB steps. In addition to measuring correct percentage of repeated sentences as the performance, listening effort was measure subjectively (by a self-report scale) and objectively (by pupillometry). The results showed decreasing SNR led to decreased performance and increased subjective effort ratings, but for pupillometry it revealed an inverted U-shaped pattern. The inverted U phenomenon suggested the participants started to disengage in the lowest SNR (i.e., the most demanding condition) and thus highest effort was spent at the medium SNR. The inconsistencies between subjective and objective measures of listening effort may not be due to the superiority of one type of measurement to the other, but rather an indication that they do not capture the same cognitive processing. In **Chapter 3**, we investigate how subjective and objective measures of listening effort can differ from each other and how we may be able to interpret each one.

1.3.2 Objective measures

1.3.2.1 Behavioural

There are several behavioural methods that can be used for evaluation of listening effort (see McGarrigle et al., 2014), such as reaction time (e.g., Houben et al., 2013). The idea behind using reaction time is that if the task is more effortful, then it should take the listener longer to perform a competing task. Alternatively, dual task paradigms can be also used to evaluate listening effort objectively (e.g., Sarampalis et al., 2009). Dual task measures are based on Kahneman's Capacity Model of Attention (Kahneman, 1973): if more attention is devoted to a task due to its difficulty, then less attention resources are available for a simultaneous task. Therefore, the reduced allocated attention for the secondary task can negatively impact the performance which can be observed via higher reaction time.

As an example for a single-task paradigm, Houben et al., (2013) used reaction time for identification of the final digit of a triplet, and the summation of the initial and final digit. Using four different SNRs (-6, -1, +4 dB, quiet), they observed reaction time increased with lower SNR in both tasks. Therefore, increasing the difficulty of the task increased reaction time which in return could be used as a sign of listening effort (Houben et al., 2013). Similar results were also found using a dual-task paradigm in a study by Sarampalis et al., (2009). Hearing-impaired participants repeated sentences presented in quiet and four-talker babble over headphones while simultaneously responding to a visual task, measured by participant's reaction time. Two different hearing aids settings (with or without noise reduction scheme) with four different SNRs [-6, -2, +2 and +∞ dB (quiet)] for the speech recognition task were used. The results of the reaction time of the secondary visual task showed an interaction between SNR and noise reduction scheme. Based on such an objective method, the authors concluded that noise reduction can decrease listening effort in the lowest SNR condition, but not in the highest SNR condition.

1.3.2.2 Physiological

While behavioural data provides us with valuable information, there are some benefits of using physiological recordings over behavioural data as an objective measure. Firstly, most of the physiological measures are time-series within an individual trial. While behavioural data can show trends over trials during an experiment, time-series data can give us a better idea of when listening effort starts to kick in or how it evolves over time in orders of milliseconds or seconds after the onset of the stimuli (Winn et al., 2018). Secondly, physiological measurements can be used in devices and technologies for the automatic evaluation of listening. While behavioural data needs to be collected and interpreted by an operator, an automated and supervised device which is trained by machine learning algorithms can record physiological data by its sensors, classify it, and make real-time decisions and adjustments without the need for listener intervention. Thirdly, as mentioned in **Section 1.1**, there are changes in listening which are not behaviourally manifest and can be detected by physiological measures. This is especially true when a wide range of task demand is used. In extreme task-demand situations (either very low or very high), physiological measures can show similar measured values due to inverted-U phenomenon. While the non-linearity of physiological effort can be a problem for any automated system to detect whether the task is too easy or too difficult, other simultaneous environmental measures by a hearing aid (e.g., sound pressure level) can be taken into account for more accurate decision making.

Physiological measures of listening effort mostly record activities in central (CNS) and autonomic nervous system (ANS) as their activities are altered by spending more effort. The most used modalities for this purpose are EEG, functional magnetic resonance imaging (fMRI), pupillometry, cardiovascular measures and skin conductance. I introduce each one briefly in the next section, but as EEG is the main outcome measure in the current thesis, it will be described in depth in **Section 1.4**.

1.3.2.2.1 Pupillometry

Pupil dilation has been associated with arousal (Darwin, 1872) and resource allocation (Granholm et al., 1996) and is caused by the interaction of the sympathetic (SNS) and parasympathetic nervous systems (PNS) (Zekveld et al., 2018). Two most used indices of listening effort in pupillometry studies are peak pupil dilation (PPD) and mean pupil dilation (MPD) which are the basic measures of task-evoked pupil response (TEPR). It has been shown that PPD and MPD are increased when more effort is invested in the task (e.g., Ohlenforst et al., 2018; Wendt et al., 2018; Zekveld et al., 2011). In these studies, often PPD and MPD started to decrease in very demanding listening conditions which may be due to disengagement from the task. However, changes in pupil size may not be only affected by listening effort, as it is also sensitive to other personal and environmental factors such as age, sound features or luminance level. While some of them may reveal valuable information, they can complicate interpretation of pupil results (Naylor et al., 2018).

The inverted U-shaped pattern of effort has been observed in pupillometry studies numerous times (e.g., Cabestrero et al., 2009; Granholm et al., 1996; Ohlenforst et al., 2017, 2018; Wang et al., 2018; Wendt et al., 2018; Zekveld & Kramer, 2014). For example, Wendt et al., (2018) showed that peak pupil dilation was altered after exposing the participants to a wide range of SNRs (12 different SNRs, varying from -20 to +8 dB) during a speech-in-noise task. Their results indicated that there was an inverted U-shaped pattern of pupil dilation that may reflect listening effort as a function of SNR. The peak pupil dilation was highest around 50% sentence recognition, decreasing when the task became easier (higher SNR) or more difficult (lower SNR). Using pupil data, the same authors investigated if different types of maskers can affect listening effort. To test this, they used two levels of SRTs (50% and 84%) in a speech-in-noise task in presence of 1-talker masker, 4-talker masker, and fluctuating noise. They observed that listening with 1-talker masker in the background increased pupil dilation slightly more compared to 4-talker masker and fluctuating noise. Although this may have indicated that 1-talker masker is more distracting compared to more energetic masking types, but comparisons of pupil dilations were drawn between listening situations that differed in SNRs (almost as high as 13 dB). Such a huge variation in SNR can lead to different pupil dilation that might have been unrelated to listening effort.

1.3.2.2.2 fMRI

fMRI is a neuroimaging technique that measures blood oxygen level dependent (BOLD) that has fine spatial resolution but suffers from poor temporal resolution (Glover, 2011). The high spatial resolution of fMRI has proved to be useful in studying brain functions during effortful listening (e.g., Kuchinsky et al., 2015; Rosemann &

Thiel, 2020; Zekveld et al., 2014) . However, there are downsides to using fMRI in hearing studies compared to other physiological measurement devices. The gradient-induced vibration in fMRI can cause interfering acoustic noise as a form of energetic masking in hearing tests (Peelle, 2014). Another issue is fMRI complexity of use compared to pupillometry or EEG, which has made it a far less desirable target for ambulatory or portable measurements of any sorts.

Despite its disadvantages, fMRI has helped hearing sciences immensely in better understanding of effortful listening. Investigating complex brain concepts such as top-down/bottom-up[1] processing of the brain during listening is more reliable with fMRI than other recordings including EEG. As an example, Davis et al., (2011) examined whether semantic context in different SNRs can lead to top-down processing in the brain. For this purpose, participants were presented with coherent (e.g., *"her new skirt was made of denim"*) and incoherent sentences (e.g., *"her good slope was done in carrot"*). The goal was to look for difference in BOLD recording when there was context to the sentence compared to when there was not (in other words, whether BOLD activities show top-down processing). Despite not finding enough neural evidence for top-down processing during perception of degraded speech in this study, one of the findings indicated that BOLD activities in left inferior frontal gyrus changed in inverted U-shaped pattern as a function of SNR (1 dB incremental increase from -5 to 0 dB, in addition to clear speech), which might have reflected listening effort.

1.3.2.2.3 Cardiovascular and skin conductance measures

Similar to pupillometry, cardiovascular measurements have also been used to evaluate SNS and PNS activity in listening task (Sherwood et al., 1990). Two main methods in cardiovascular studies of listening effort are heart rate variability (HRV) to measure PNS (e.g., Cvijanović et al., 2017; Mackersie et al., 2015) and pre-ejection period (PEP) to measure SNS (e.g., Plain et al., 2020; Richter, 2016). In these studies, decrease in HRV and PEP reactivity are considered as markers of increased SNS and PNS activity during effortful listening.

Another method for measuring SNS activity is skin conductance which reflects the amount of moisture excreted from the eccrine glands. Increase in skin conductance is

[1] For a brief introduction of top-down and bottom-up concepts, see **Section 1.4.4.4**.

marker for increased SNS activity which may be as a result of increased task difficulty (Mackersie et al., 2015).

In the study by Mackersie & Calderon-Moultrie (2016), both HRV and skin conductance were used simultaneously to look for changes in listening effort. Participants were asked to repeat the sentences they heard in a speech-in-noise paradigm. The sentences were presented either at normal speaking rate at 0 dB SNR or a fast-speaking rate at +3 dB SNR, both conditions approximating 80% correct word repetition. The results showed that HRV reactivity for high frequencies decreased, and skin conductance reactivity increased during the more effortful fast-rate condition compared to the normal-rate condition. While the authors concluded HRV and skin conductance can be used as indices for listening effort, they also mentioned the possibility of these being measures of stress and/or motivation.

1.4 EEG

1.4.1 Physiology

EEG measures electrical activities of the brain by placing a cap with attached electrodes on the scalp. It was Hans Berger who showed this for the first time by recording electrical activity of a human’s brain in changes of voltage over time (Berger, 1929). The electrical activity recorded by EEG is the summation of postsynaptic potential in the brain that typically lasts over tens or hundreds of milliseconds. When neurotransmitters bind to receptors on the membrane of the postsynaptic cell, ion channels are altered. This leads to a voltage change across the cell membrane called postsynaptic potential. EEG signals often range in microvolts, but the amplitude of the signals depends on different factors such as thickness of person’s scalp or skull or cortical folding patterns (Luck, 2014).

1.4.2 Brain oscillations

After the discovery by Berger, Adrian and Matthews discussed the findings and introduced the concept of oscillatory waves in the brain. They named the most dominant oscillations in the brain within 10-12 Hz frequency “alpha rhythm” (Adrian & Matthews, 1934). Since then, EEG oscillations have been key features in many neuroscientific studies.

Oscillations within different frequency bands are categorized into distinct brain waves. Any cognitive task requires certain oscillations in one or several specific bands in certain areas of the brain. Those associated oscillations can either be used as neural markers of a cognitive processing or as a way of understanding how the brain functions in certain situations.

Slowest waves (0.5-4 Hz) are called delta band which is seen in deep sleep and is prominent in the frontocentral regions. Slightly faster waves (4-8 Hz) are called theta band which is also most prominent in the frontocentral region of the brain and can coordinate memory processing (Kragel et al., 2020) or encode new information (Hasselmo & Stern, 2014). Probably the most studied band among EEG oscillations is alpha waves (8-12 Hz) which manifests during resting state or also during several different cognitive tasks such as selective attention (Foxe & Snyder, 2011) or increased memory load (Jensen et al., 2002). Alpha band is mostly present in posterior regions of the brain. Faster oscillating waves compared to alpha are called beta band (12-30 Hz) which is involved in somatosensory processing and motor control (Barone & Rossiter, 2021) and gamma band (30-100 Hz) which are thought to integrate the processing of neuronal networks in the brain (Kaiser & Lutzenberger, 2003) .

However, there are issues with such traditional labels with fixed definitions. The first issue is that even within a defined band, there can be two functionally independent oscillations. For example, alpha power can be divided into low (8-10 Hz) and high (10-12) alpha power. It has been shown that low and high alpha have separate functions which each may reflect attentional process and stimulus-related cognitive process respectively (Klimesch et al., 1993).

The second issue is that the frequency range of these oscillations has changed often in the literature as they have been sensitive to diverse experimental manipulations. Also, other physiological factors such as age (e.g., Rondina et al., 2019) can change the range and power of such oscillations. Even within an individual alpha peak frequency is liable to rapid changes at shorter time scales (e.g., Mierau et al., 2017). For these reasons, it has been argued that such traditional labels may have outlived their usefulness (see Weisz & Obleser, 2014). The frequency range of such brain waves should not be strictly fixed, but rather examined carefully within a particular experimental design and set of participants. This issue with fixed EEG bands is carefully examined throughout all the chapters of this thesis. In **Chapter 2** and

Chapter 3, divided alpha power (low vs high) will be further explored to see if they differ from each other.

1.4.2.1 Synchronization and desynchronization

To measure the power or phase of brain oscillations, different mathematical methods such as a Fourier or wavelet transform are used. The essence of these methods is that any waveform can be deconstructed into a set of sine waves (of varying intensity and phase) or wavelets, respectively. This does not mean, however, that the waveform truly consists of a set of sine waves. Hence, when applying the transforms to an EEG signal, power at a given frequency does not mean that the brain is oscillating at that frequency. Therefore, obtaining non-zero values for power calculations is mathematically inevitable and not an evidence for physiological oscillation (Luck, 2014). In addition to that, absolute values of brain oscillations are prone to unwanted variance over experiments (e.g., alpha increase due to fatigue) or participants (e.g., some individuals may have thicker scalp and thus lower magnitude signals) that may complicate the interpretation of the results. In order to claim that a given value in power is actually reflecting oscillation, power bands need to be normalized to a task-free baseline. The relative changes which are discriminable in the power spectrum (for example in the form of a peak) can be considered as neural oscillations of the brain. The changes of power compared to the baseline can be positive or negative. If the changes are positive, the power band is considered to be synchronized due to the stimulus (i.e., event-related synchronization; ERS). Conversely, if the changes are negative, the oscillation is desynchronized (i.e., event-related desynchronization; ERD).It is also important to notice that ERS and ERD refer to relative measures and are dependent on the baseline. Baseline can be chosen during a silent period or during a secondary task such as reading an instruction or listening to background noise. Determining whether the task or baseline is passive or active (and if so, what kind of task) can also change the outcome of ERS/ERD (Weisz & Obleser, 2014). As an example, it has been observed that with passive baseline (listening to background noise), alpha ERS is significantly different during active listening where participants listened to a speech intently compared to passive listening where the main task was visual and the same speech is ignored by the participants (Dimitrijevic et al., 2019). Therefore, any interpretation on the frequency domain of EEG signals should be drawn with great cautious.

1.4.3 Working memory

Effortful listening requires allocation of cognitive resources in the brain. By investigating different tasks, such as working memory tasks which can vary the allocation of resources, changes in listening effort can be studied. Working memory is the process which is involved in the control and regulation of information in order to complete a complex cognitive task (Miyake & Shah, 1999). In working memory terminology, most memory tasks consist of three key phases: encoding, maintenance and recall (Pinal et al., 2014). In a listening task, the "encoding" phase involves listening to the signal whether in the presence or absence of distractors. The "maintenance" phase comes after the signal of interest is finished and requires keeping the information in short term memory (STM). It is important to note that the terms "working memory" and "STM" are often used interchangeably. However, in this thesis, as suggested by Baddeley (2012), temporary storage of information will be referred as STM, and a combination of storage and different phases of listening as working memory. Finally, the "recall" phase requires using the information stored in the STM, in order to interpret the signal to complete a task.

However, often encoding and maintenance are entangled, and their neural processing may overlap. Information that is maintained in working memory can be updated with new encoded information. Therefore, for successful working memory performance, information should be replaced and updated when necessary. The process of constantly manipulating information and maintaining it is an important feature of working memory which can happen in a wide range of complex cognitive task such listening to a continuous speech. The overlap between encoding and maintenance in working memory can be done without much manipulation operations if the task is not demanding. However, in demanding tasks, relevant manipulations are triggered by task demand which manifests in activation of certain brain networks (Nyberg & Eriksson, 2016) which can be visible through different neural correlates such as power changes in different bands.

1.4.4 Alpha power as a measure of effort

The power of alpha oscillations is a commonly used neural correlate of listening effort.

There are several theories on the role of alpha power during a cognitive task. In one of the earlier theories, Pfurtscheller (1996, 2001) introduced “cortical idling”. With this theory, he argued that alpha ERD reflects an activated level of cortical neurons which can be interpreted as increased cortical excitability. If alpha ERD is characteristic for an activated cortical area, then alpha ERS may describe a resting or cortical idling in which no information is being processed (Pfurtscheller, 2001; Pfurtscheller et al., 1996). Later, Jensen and Mazaheri (2010) argued that cortical idling does not explain why an improvement in behavioural performance is accompanied by an increase in alpha power. Instead, they proposed “functional inhibition” theory can cover this shortcoming. The idea of this theory is that information is routed by inhibiting task-irrelevant pathways which is reflected by increase in alpha activity. This is consistent with the notion that optimal task performance is dependent on inhibition of task-irrelevant regions in order to allocate resources to task-relevant regions (Jensen & Mazaheri, 2010).

Klimesch and colleagues (2007, 2012) also argued that alpha ERS has a more complex role than what Pfurtscheller (1996, 2001) previously suggested. Based on many experiments that have shown alpha ERS reflects “functional inhibition”, he suggested alpha ERS is linked to both timing and blocking of information processing. Based on these attributes, alpha could be representing attention, as attention enhances the processing of information by blocking irrelevant processing within a specific time (Klimesch, 2012). With external information, this process is usually under top-down control which enables “the ability to be consciously oriented in time, space, and with respect to the meaning of all entities surrounding the individual” (Klimesch, 2012, P.612).

In **Section 1.2**, it was pointed out that difficult listening situations can result in noisy STM representation which leads to mismatches between the representation of words in episodic and semantic memory. The working memory should engage actively in demanding situations in order to disambiguate of what had been heard (Rönnberg et al., 2013). Alpha oscillations as means of “functional inhibition” is one of the mechanisms that plays an important role to overcome the mismatch, as it facilitates the allocation of resources to task-relevant regions. The higher demand on working memory requires more inhibition to “shut down” task-irrelevant regions of the brain and thus more alpha power is generated in those regions (Wilsch & Obleser, 2016). It should be noted that changes in alpha power are not specific to auditory tasks, as it has been observed during non-auditory tasks that alpha power increases with higher demand (e.g., Bonnefond & Jensen, 2012; Jensen et al., 2002; Manza et al., 2014; Tuladhar et al., 2007).

1.4.4.1 Effort during maintenance

Changes of listening effort are not limited to the encoding phase (i.e., listening), but also the maintenance phase of an auditory working memory task. In a study by Obleser et al. (2012), normal-hearing participants performed an auditory working memory task which involved listening to and remembering digits with three different memory load (2, 4 and 6 items) and three different acoustic degradation (16, 8 and 4 bands in noise vocoding of the items). The authors specifically investigated the maintenance phase after listening to speech with alpha power to predict response time as an index for listening effort. They found out the degraded speech (a perceptual challenge) and increased memory load (a capacity challenge) increased central–parietal alpha activity during that maintenance phase and increased response time. The alpha increase was superadditive when most degraded was combined with highest memory load. The conclusion was that as more degraded speech and increased memory load led to increased alpha and response time, then both perceptual and capacity challenges require increased "functional inhibition".

The notion that increased alpha power during maintenance can be a sign of listening effort was also suggested in Wisniewski et al. (2017). In this study, two frequency-modulated sweep tones (A and B) were presented back-to-back. The task for the participants was to identify tone A or B after listening to them. One of the tones always had a standard rate and the other one had a varying rate which made the task easy, difficult or impossible. After a maintenance period, participants were presented with the tone. They observed that the modulation of alpha and theta power was highest in difficult condition during maintenance, but lower for easy or impossible listening situations. This inverted U-shaped pattern in maintenance can be another evidence that alpha power reflects effort. In another study by Wisniewski et al., (2015), a speech-in-noise task was used to evaluate the changes of EEG power during maintenance across five different SNRs. They used independent component analysis (ICA) to show that during maintenance, although activity increased in the theta and low beta bands (6 and

14 Hz peaks), and marginally decreased in the alpha band (centered at10 Hz), there were no effects of SNR. The authors concluded these EEG changes showed task engagement more than stimulus features.

Increased activity of alpha/theta power during the maintenance phase is not limited to auditory studies. Increased alpha activity during maintenance with higher memory load has been observed in several non-auditory working memory tasks (e.g., Jensen et al., 2002; Tuladhar et al., 2007). This suggests that alpha activity during maintenance is not auditory-specific and has more global functions such as processing stored information in memory.

1.4.4.2 Alpha comparison to subjective effort

In **Section 1.3.1**, it was discussed how objective measures can deviate from the subjective measures of listening effort. This is no exception for EEG studies. There are several EEG studies that have opposing views on this matter. For example, a study by Decruy et al., (2020) showed that self-report scales of listening effort and EEG alpha power do not follow the same pattern. In this study, 40 sentences were concatenated together with a short silent gap between them. Six SNRs that corresponded to 20%, 35%, 50%, 65%, 80%, 95% word recognition for each individual were used, along with speech in quiet, to manipulate task demand. While self-report showed negative linear relation with SNR, alpha power showed inverted U-shaped pattern with 60-80% speech intelligibility being the highest alpha results.

However, there are also experiments that have shown alpha activity is correlated to self-reported listening effort. In an experiment by Wöstmann et al., (2015), the modulation of alpha power in young and elderly participants were compared together. Participants were instructed to listen to two spoken digits in presence of a distracting talker and choose whether the second digit was lesser or greater than the first digit. Acoustic detail (temporal fine structure) and predictiveness (guessing if the second digit is smaller or larger) were changed to see their effects on alpha power. The results showed that decreasing acoustic detail and predictiveness (i.e., higher task demand) led to increased alpha power. Acoustic detail had a stronger effect on elderly, suggesting that alpha activity might be changing with age. Participants were also asked about their subjective listening effort (post experiment) and their confidence in providing the right answers (post trial). The results showed that alpha power was positively correlated to the subjective effort and negatively correlated to confidence levels. A possible reason for the significant linear correlations between alpha activity and self-reported effort and confidence in this experiment was the small range of task demands that was not wide enough (task performance was designed to range 60-80% accuracy) to cover the disengaged portion of the listening effort curve due to very high task demand.

In an exploratory attempt to investigate different brain regions and bands which are correlated to subjective rating of listening effort in cochlear implant (CI) users, Dimitrijevic et al., (2019) used source localizing technique on EEG data. Participants listened to a digit-in-noise recognition task in both passive and active conditions. Active listening required identification of the presented digit which followed one-up one-down procedure to yield 50% correct identification. For the passive listening condition participants watched a movie and ignored the digits. The results showed that alpha power in the left inferior frontal cortex was highly positively correlated with the subjective rating of effort. However, using two levels of demand would have not been enough to capture the non-monotonic relationship of listening effort and demand. Similarly in the study by Wöstmann et al., (2015), the range of task demand was not broad enough (despite using four levels of task demand) to reveal any non-monotonic relationships between listening effort and demand. It will be shown in **Chapter 3** that when task demand varies from very low to very high during a speech-in-noise task, then subjective and objective measures of effort do not correlate to each other.

1.4.4.3 Alpha correlation to speech intelligibility

Despite large number of reports on alpha power relating to listening effort, there are also other studies that could observe alpha association to speech intelligibility and not effort, especially during alpha ERD (e..g, Dimitrijevic et al., 2017; Obleser & Weisz, 2012).

The evidence for alpha correlation to speech-related performance was observed during an experiment in which participants were asked to listen to digits either actively or passively while watching a movie. The speech consisted of monosyllabic digits with a +2 dB fixed SNR (close to 100% digit-recognition accuracy) and a fixed SNR at the SRT (close to 50% accuracy). The authors observed that active listening produced alpha ERS in some listeners, whereas passive listening produced almost no such oscillation. Using source localization, they showed that digit recognition performance was related to alpha ERD in temporal lobe but not to ERS observed in central/parietal regions (Dimitrijevic et al., 2017).

Similar conclusion was also drawn in a study by Obleser & Weisz, (2012) using subjective speech comprehension ratings. Four levels of spectral (2, 4, 8, 16 band vocoding) and four levels of envelope degradation (2, 4, 8, 16 Hz envelope) during presentation of a single word were used to manipulate task demand. The grand average results showed that from 500 ms post word onset alpha ERD occurred. Using subjective comprehension ratings, they showed that there is a negative correlation between alpha and speech comprehension (i.e., more alpha ERD led to better speech comprehension). The authors concluded that alpha ERD in the beginning of the speech reflects speech intelligibility. They also showed that less acoustic detail (i.e., more demanding) led to increased alpha power, which is in line with theories about alpha and listening effort (i.e., more effort leads to more alpha). Therefore, the question remains whether alpha power is reflecting speech intelligibly or listening effort.

It seems unlikely that alpha power can reflect both speech intelligibility and listening effort simultaneously, at least when task demand covers a wide range. As mentioned in **Section 1.2**, we know that listening effort is not always increased with increased task demand. On the other hand, speech intelligibility is decrease with increased task demand. Therefore, it is highly unlikely one measure can reflect both at the same time. Having said that, alpha might reflect *either* of them depending on the task. We will tackle this issue in Chapters 3, 4 and 5.

1.4.4.4 Top-down or bottom-up

Irrespective of whether alpha activity reflects listening effort or speech intelligibility, it is not entirely clear whether alpha activity is a result of top-down or bottom-up attention in the brain. While top-down attention describes voluntary selection of listening to a stimulus, bottom-up attention is exogenous and sensory-driven by a stimulus due to its saliency (for review see Katsuki & Constantinidis, 2014). Changes to alpha power due to listening effort (or speech intelligibility) can be caused by either top-down or bottom-up processing. However, most of the theories on alpha power mainly associate it to top-down attention (Jensen & Mazaheri, 2010; Klimesch et al., 2007)

One EEG study by Wöstmann et al., (2017), specifically investigated whether alpha power reflects top-down or bottom-up attention during effortful listening. In this study, participants listened to a serial order of 9 to-be-recalled digits which lasted approximately for 6.5 s. This was followed by a 5-s maintenance phase during which one to-be-ignored distractor sentence was presented. Distractibility of the to-be-ignored sentence was manipulated by noise-vocoding (1, 4 and 32 channels). They showed that in the maintenance phase, higher levels of acoustic details in the distracting speech disrupted listeners' serial memory recall and increased their alpha power activity. The authors concluded that if the increase in alpha power in previous studies (e.g., McMahon et al., (2016); Obleser et al., (2012)) reflected higher acoustic detail of the target speech, then higher acoustic detail of distracting speech should have increased the alpha power. Since this did not happen, alpha activity probably did not reflect acoustic details (i.e., bottom-up), and rather reflected the amount of effort spent on the listening task (i.e., top-down) which increased when the task was more difficult, regardless of changing acoustic details in target or distracting speech. This study provided evidence that alpha activity during effortful listening is a proxy of top-down control. It can be argued that top-down alpha power serves as an interface with the external world and is related to perception (Klimesch, 2012).

1.4.4.5 Alpha in hearing disorders

In **Section 1.4**, it was mentioned that listening successfully depends on engaging task relevant and disengaging task-irrelevant brain areas which are manifested through alpha power activities. For this reason, decrease in ongoing alpha responses may be a key feature for many hearing-related problems as it reflects disturbances to the excitation and inhibition of different brain regions (Weisz & Obleser, 2014).

Listening disengagement can also happen earlier with increased task demand in hearing impaired. To test this idea, Petersen et al., (2015) recruited three groups of participants that were fitted with hearing aids: no, mild and moderate hearing loss. In an auditory working memory task, participants were presented with different memory load (2, 4, or 6 digits to be remembered) embedded in a background noise level (4 dB, 0 dB, or -4 dB relative to the individual level at which 80% of the words were correctly recalled in noise) during the encoding phase. After the stimulus-free maintenance phase, participants indicated if a digit appeared in the sequence of presented digits in the recall phase. The effects of increasing memory load and background noise level on alpha activity during the delay were modulated by the degree of hearing loss. That is, participants suffering from a higher degree of hearing loss experienced disengagement with increasing task difficulty (i.e., inverted U in alpha activity), which was not observed for the participants with mild or no hearing loss. They suggested that more severe hearing loss can cause neural activity breakdown as a result of spending more resources than people with better hearing.

An interesting example of the role of alpha power in hearing disorders can also be seen in tinnitus. Weisz et al., (2005) showed that tinnitus patients have significantly less alpha activity in resting state. To evaluate if, in fact, alpha power is related to any tinnitus distress, Hartmann et al., (2014) exposed tinnitus patients to three different treatments: Neurofeedback[1], repetitive transcranial magnetic stimulation (rTMS)[2] and sham in which the same parameters as rTMS were applied, but the coil was tilted 45°. The results showed that the highest increase in alpha power, and highest decrease in tinnitus distress followed neurofeedback treatment (Hartmann et al., 2014). While speculative and without a control group for neurofeedback treatment, the decrease in tinnitus distress level after neurofeedback session might reflect that observed alpha ERS in this treatment follows the “functional inhibition” theory. Increased alpha via neurofeedback treatment may have led to better inhibition of task-irrelevant areas, and thus followed by decrease in tinnitus distress. Therefore, not only can alpha power be a neural correlate for hearing disorders and some of their consequences, such as listening effort, but it may also be useful in clinical treatments.

1.4.4.6 Conflicting alpha patterns

There are a number of contradicting studies that have shown alpha power decreases with task difficulty. For example, EEG and pupillometry were simultaneously used in a sentence recognition task (lasting ~5 s) on young normal-hearing participants at each individual’s 50% and 80% SRT with two different levels of spectral degradation (6

---

[1] Neurofeedback is a method in which a representation of brain activity is shown to the user in real-time to help them in self regulate their brain activity (Enriquez-Geppert et al., 2017).

[2] rTMS is a therapeutic approach that stimulates a small current into the cortex. This current causes depolarization and hyperpolarization of the neurons triggering neuronal activation (Zuchowicz et al., 2019).

and 16 channel noise vocoding). While, similar to the literature discussed in **section 1.3.2.2.1**, mean pupil dilation increased in the more spectrally degraded condition (6 channel), alpha power decreased in this condition. No effect of SNR was observed on either pupil or EEG data (Miles et al., 2017). In a similar design, the same research group used a range of fixed SNRs (-7 to +7 dB in 1 dB increments; a total of 15 levels to cover a full range of 0-100% in speech intelligibility) with 6-channel and 16-channel speech degradation. Again, there was increased pupil dilation and decreased alpha power in the more spectrally degraded condition, averaged across SNRs. However, in the 16-channel condition, decreasing SNR led to increased alpha power. These conflicting results across literature might be a warning sign that the pattern of alpha ERS/ERD goes beyond an abstract concept like listening effort.

Another study that showed alpha power decreased with higher task demand was the study by Hauswald et al., (2020) which used MEG on two experiments. In both experiments, participants listened to a continuous speech (approximately lasted between 30 s – 3 mins) and were asked to choose from two nouns that had occurred within the last four words of the trial. In experiment 1, three different levels of noise vocoding (original, 7 and 3 channels) and in experiment 2, three additional vocoding levels (5-channel, 2-channel and 1-channel) were also implemented. Both experiments showed that performance and alpha power decreased with more degradation of the speech. However, both studies showed that low-frequency oscillations (1–7 Hz) in frontal regions showed an inverted U-shaped pattern for speech tracking (i.e., coherence between speech envelope and brain activity) with changing degradation levels.

1.4.4.7 Alpha phase

Given that EEG data is rich in information, other methods can be used to extract features in EEG for evaluation of listening effort. For example, Bernarding et al., (2017) proposed a novel measure to look for changes in listening effort in hearing aid users; the novel measure was the instantaneous phase information of the ongoing EEG activity by the wavelet transform corresponding to a pseudo frequency of 7.68 Hz (low alpha). They postulated that the phase should be more uniformly distributed on the unit circle[1] for the less demanding listening condition. To compare the results of the proposed method with a more traditional method, self-reported listening effort and subjective intelligibility were also collected. In the experiment, four different hearing aid settings (strong/medium/no directional speech enhancement and omnidirectional microphone setting) were used during presentation of short sentences and a 10-minute-long story. The task was to repeat the words for the short-sentence stimuli and answer a question relating to the story. The results of both subjective and objective measurements showed significant decrease in effort in the directional microphone settings compared to the omnidirectional microphone, but no significant changes among any of the three directional settings. In terms of low alpha phase, they concluded that the distribution of the instantaneous phase of low alpha band reflects cognitive effort and is more clustered during demanding listening situations. However, because low alpha phase was not directly compared to (low/high) alpha power, it remained unknown that which EEG measure can be more sensitive to the changes of listening effort in different levels of task demand.

1.4.4.8 Pre-stimulus alpha

Alpha power activity prior to stimulus presentation may also reflect the success chance of listening. In a study by Alhanbali et al., (2021), in an auditory working memory task, participants listened to 6 single digits and memorised them during a maintenance phase. The task was for the participants to recognize if a probe digit was presented during the initial sequence or not. Alpha power in the maintenance phase, as well as alpha power in the pre-stimulus phase were positively correlated to participants' performance in remembering the digits. The increase in pre-stimulus alpha power might have been an indicator of increased auditory attention that could have suppressed any unwanted distraction during the task. This interpretation could be in line with the study by Wöstmann et al., (2015) that decreasing stimuli predictiveness would increase alpha power due to increased auditory attention. However, any

[1] Unit circle is a way of quantifying the synchrony phase of specific EEG oscillation. It shows the distribution of the instantaneous phase which was extracted by applying a complex continuous wavelet transform in the mentioned study (Bernarding et al., 2017).

interpretation on pre-stimulus alpha has been based on absolute values of power which has some critical shortcomings as discussed in **Section 1.4.2.1**.

1.4.5 Measuring effort with theta

While not as frequent as alpha power, theta power has also been used as a neural correlate of listening effort in auditory and non-auditory studies. More specifically, frontal midline theta can be measured during attention or working memory tasks. Similar to alpha, increased task difficulty leads to increased frontal theta activity (Onton et al., 2005).

In **Section 1.4.4.3**, the possibility of alpha activity showing different cognitive aspects in different auditory tasks was discussed. Similarly, listening effort might be reflected in other measures such as theta ERS during specific auditory tasks as well. As an example, in a delayed pitch discrimination task, participants were instructed to recognize the interval containing higher pitch in two square wave stimuli differing in pitch. In a “Roving” condition, the lowest pitch stimulus was randomly selected on each trial (from 840 to 1160 Hz). In a “Fixed” condition, the lowest pitch was always 979 Hz. The difference between the Fixed and Roving condition was that in Fixed condition participants could respond immediately following the first stimulus while in the Roving condition they needed maintenance of the first tone for comparison to the second. The results showed that while alpha power was unchanged, frontal midline theta in Roving was significantly increased in comparison to Fixed (i.e., when there was a need for maintenance). The authors concluded that not all difficult listening tasks will be accompanied by the same EEG indices of listening effort (e.g., alpha or theta ERS) (Wisniewski et al., 2018).

In fact, using theta power as an objective measure of listening effort has been shown to be correlated to subjective ratings of effort. In the study by Wisniewski et al., (2015), speech-in-noise (~3 s) were presented to the participants with four different SNRs (-12, -6, 0, 6 dB). The task included a baseline period with background noise, listening phase and maintenance phase which was without any background noise. EEG theta power and self-report scale were used to assess listening effort. Using ICA, medial frontal components showed increasing theta power with decreasing SNRs (i.e., increased task demand) during the listening phase. Self-report scales also were positively correlated to theta power during listening. Based on the previous literature on listening effort, it might be expected such a wide range of SNR (18 dB wide) shows an inverted U-shaped pattern. Instead, theta power was linearly changed with SNR (similar to subjective ratings). So, it is plausible the changes of theta power were not reflecting listening effort, but rather showing the changes of SNR or sound pressure level (SPL).

Theta phase is another measure that can be used to assess listening effort. In a standard auditory-oddball paradigm, participants were instructed to discriminate between deviant high-frequency tones (10%) interspersed among low-frequency ones (90%) “near” or “far” separated in frequency. Using inter-trial phase coherence (ITPC), the near condition (i.e., more difficult) led to great frontal theta and gamma compared to the far condition. However, this only happened during active listening and presenting the participants with the same stimuli during passive listening, no such results were observed. The conclusion was that ITPC differences may reflect differences in attention-modulated stimulus encoding (Wisniewski, 2017).

There are also contradicting results on theta power as a measure of listening effort. For example, Marsella et al. (2017) demonstrated that increasing demand in a speech recognition task consisting of disyllabic words presented free-field from a front loudspeaker in children with asymmetric sensorineural hearing loss did not increase frontal theta. There were four noise conditions (in order of demand): no noise (in quiet), four-talker babble presented from one loudspeaker 90° to the worse-ear side, babble noise presented from two loudspeakers at ±45°, and babble noise presented from one loudspeaker 90° to the better-ear side. While they did not observe any significant change in theta activity, there was an inverted U-shaped pattern in alpha power. Alpha was increased in intermediate difficulties (binaural noise and noise to the worse ear) compared to the quiet condition, but it also decreased in the most demanding listening condition (noise to the better ear).

1.5 Thesis objectives

The goal of the thesis is to investigate how listening effort can be measured objectively using EEG in ecologically valid situations. The focus of the EEG analysis will be on alpha ERS/ERD, but theta and beta bands will be investigated in an exploratory way as well. Five different experiments will be reported and discussed through chapters 2 to 5, which all include speech-in-noise task in demanding situations.

In **Chapter 1**, the concept of listening effort was introduced and how different personal (e.g., motivation) or environmental (e.g., speech degradation) factors can manipulate effort. Subjective and objective measures of listening effort were also reviewed and how EEG signals could be useful to evaluate listening effort objectively. In **Chapter 2**, the concept of motivation will be discussed more and how monetary reward can be used to manipulate motivation in a task. Task demand will also be varied by changing SNR to see how different levels of task demand can influence one's motivation to invest effort.

In **Chapter 3**, reverberation will be introduced and how listening can be affected when listener is in different acoustic environments. For this purpose, different simulated rooms in an anechoic chamber, each with distinct reverberation time, will be used. By manipulating task demand with SNR, the interaction between reverberation time and SNR in each room will be explored. For this study, subjective measures of listening effort with a self-report scale will also be investigated and compared to EEG data.

In both **Chapter 2** and **Chapter 3**, the speech material are short and interrupted sentences. To move towards more ecologically valid situations, continuous speech will be presented to the participants in **Chapter 4**, as listening to a continuous speech occurs more often in daily life compared to listening to a single short sentence. The aim of this paradigm is to see if the pattern of alpha power is any different during continuous speech compared to short speech. After observing how alpha power changes with varying SNR in continuous speech paradigm, noise reduction scheme of hearing aids will be investigated to see whether it affects listening effort in hearing impaired.

In **Chapter 5**, ear-EEG will be introduced as a novel and ambulatory recording of brain signals. The usefulness of ear-EEG to measure listening effort will be evaluated (more specifically alpha power) during a continuous speech paradigm and compared to traditional (scalp) EEG.

In **Chapter 6**, the findings across these five experiments will be discussed and how they may help us understand the role of alpha power during effortful listening in a real-life setting.

Chapter 2
The role of motivation in shaping listening effort

2.1 Introduction

Listening effort can be modified by personal (such as motivation) or environmental (such as task demand) factors. In this chapter it will be investigated that how the interaction of both factors should be considered when studying listening effort.

2.1.1 Listening motivation

As mentioned in **Chapter 1**, studies which have used physiological measures of listening effort have often found an inverted U-shaped pattern of listening effort (e.g., Wu et al., 2016; Wisniewski et al., 2017; Wendt et al., 2018; Decruy et al., 2020), meaning that people increase effort as demand increases, but only up to a point. This phenomenon is usually seen when the range of task demands varies from very low to very high and forms an inverted U-shaped pattern of listening effort. In **Section 1.2**, the role of motivation in expenditure of effort was discussed in Brehm's theory of motivation intensity (Brehm & Self, 1989) which is not listening specific but a broad cognitive model. It is hypothesized that people only put in more effort for difficult tasks if their effort expenditure is perceived to yield appropriate benefit. If the task becomes too difficult, the effort will be low because its expenditure will be seen as yielding a return of insufficient value (Wright, 2008). In other words, effort decreases when the benefits of performing the task do not outweigh the costs of allocating resources to a task that is too difficult. Also, when a task is too easy, not many resources are required for optimal performance, so spending more effort does not lead to more behavioural advantages. It is in the intermediate cases that listeners apply greater effort due to the potential of optimizing performance. The importance of motivation has specifically been considered in relation to listening effort through FUEL. This framework drew upon MIT (Brehm & Self, 1989) and proposed that listening effort is not just a function of task demand (e.g., degraded speech), but also a function of the individual's motivation, whether intrinsic or extrinsic. In other words, the amount of cognitive resources one is willing to expend in a listening task at a particular level of difficulty, is also based on how motivated the person is (Pichora-Fuller et al., 2016).

The expenditure of effort is costly and the trade-off between effort and performance can be investigated through behavioural economics. This framework studies when and why people choose to listen, based on the cost and benefits of the task. The decision to spend more effort is usually made by an adaptive control in the brain to optimize performance during demanding tasks and sustain engagament. When the outcome is worth the effort, the adaptive control upregulates the neural activities for better speech understanding (Eckert et al., 2016).

While task demand is enforced by environmental factors or hearing loss, the source of motivation can be *intrinsic* as well as *extrinsic*. Intrinsic motivation is driven by one's own sake and satisfying results, whereas extrinsic motivation is formed by an outside reward or punishment. For example, a listener's intention to engage in a friendly conversation triggers an intrinsic motivation to listen. In such a scenario, there is probably no benefit in listening, but the listener is still willing to spend the effort for their own satsifaction. On the other hand, if a reward is given for successful listening (such as monetary reward which is mostly unrealistic in real life) then the motivation is extrinsic. But even then, the borderline between intrinsic and extrinsic motivation cannot be well defined because extrinsic motivation can be internalized (Hidi, 2016). In the above example, the listener does not put effort to necessarily gain monetary reward, but they might do so to gain personal satisfaction of being rewarded.

In order to manipulate motivation in the labarotary, researchers commonly rely on manipulating extrinsic motivation, often by introducing some forms of reward such as monetary incentives (e.g., Koelewijn et al., 2018; Kostandyan et al., 2019; Plain et al., 2020) . For this reason, for the rest of the chapter, whenever the word "reward" is being used it implies the manipulation of extrinsic motivation.

The theories and models on motivation and effort have been tested in several different listening studies using subjective and objective measures of effort. For example, Picou and Ricketts (2014) manipulated whether or not participants were evaluated at the end of an auditory task as a means of operationalizing motivation: participants either only listened to the speech (low motivation), or listened to the speech and answered quiz questions about it (high motivation). With two levels of SNR (targeting 50% and 80%

correct performance), they showed that in a difficult listening condition, participants reported increased effort when they were evaluated at the end of the task (high motivation) compared to when they were not (low motivation). However, for a relatively easy listening condition, there was no such effect. Despite promising results, the form of motivation used in this study was difficult to quantify, meaning that the motivation was binary; either it was there, or it was not. In real life, however, there are levels to one's motivation and is hardly binary. Another problem might be that motivation in this study may have been rewarding (taking pride in correct answer) or punishing (embarrassment of incorrect answers) for different people and thus had different affects. Also, the measured effort was subjective and there are studies that have shown subjective effort can be different than objective effort (e.g, Zekveld & Kramer, 2014). Since then, more studies have implemented other forms of extrinsic reward that are easier to quantify and compare (such as monetary reward) and more objective outcomes of effort (such as pupillometry or EEG).

One such study that used monetary reward to operationalize motivation and changes of pupil dilation as an objective measure of listening effort found no interaction between task demand and motivation (Koelewijn et al., 2018). Instead, higher motivation increased effort regardless of the difficulty of the SRT tasks (50% and 85% correct). These findings were not in total agreement with FUEL that there is an interaction between demand and motivation, but instead motivation led to increased effort regardless of task demand. One plausible explanation for this finding is that the range in task demand was not great enough, including both too-difficult and too-easy extremes, to elicit an interaction effect with reward. Furthermore, there was no improvement in behavioural performance due to the effort. The opposite outcome was perhaps expected, that if participants were to spend more effort, they should be able to perform better.

To understand how reward and effort impact performance, another study investigated them in a visual-attention task (Kostandyan et al., 2019) using pupillometry as well as simultaneous EEG to measure effort. In the first experiment, trials with monetary reward were presented within a block (block-based) which led to higher activities of EEG alpha power and pupil size (i.e., thus more effort) compared to no-reward block, but without any benefits in performance. In the second experiment, trials with and without monetary reward were presented randomly within a block (trial-based). In this experiment they observed increased effort as well as increased performance in rewarding trials. They concluded that reward-driven effort and performance may operate on different time scales and a preparatory phase is required to achieve behavioral benefits (Kostandyan et al., 2019). This might be one explanation that why Koelewijn et al., (2018), which used block-based reward design, did not observe effects of reward on performance despite changes in effort.

Other than EEG and pupillometry, cardiovascular PEP is another marker for objective effort that has been used in motivational studies. For example, Richter (2016) used PEP to observe the interaction between reward and effort in an auditory pitch discrimination task. Two levels of task demand with 3 Hz (hard) and 20 Hz (easy) pitch discrimination with two levels of monetary reward (high and low) were tested on normal hearing participants. The results showed that there was an interaction between reward and effort; when the task demand was high, participants spent more effort in high reward situation compared to when the task demand was low. However, similar to the two previous studies, Richter (2016) also did not see an effect of reward on performance with a block-based reward paradigm.

Conversely, there are studies that have reported neither main nor interaction effect of reward. For example, Plain et al., (2020) used PEP as a marker of listening effort in a speech-in-noise task. Six different SNRs varying from very easy to very difficult, with two levels of monetary reward, were tested on normal hearing participants. Even though higher reward increased the performance accuracy in the participants, but it did not influence PEP. The authors speculate that due to long sessions of listening, the assigned monetary reward were not motivating enough for the participants.

2.1.2 Objectives

The objective of this chapter is to investigate how the interaction of task demand and motivation modifies listening effort and if more effort is invested, whether it benefits performance or not. Based on the inverted U-shaped curve previously found in objective listening effort studies (e.g., Wu et al., 2016; Wendt et al., 2018), and MIT (Brehm & Self, 1989), it is hypothesized that there would be a quadratic interaction between task demand and motivation when the demands range from very easy to very hard. For this purpose, task demand and extrinsic motivation was varied by having participants listen to and repeat speech in four different SNRs (-8, -4, 0 and +4 dB) at two levels of monetary reward (0.5 DKK or ~0.06 € for low reward and 7.5 DKK or ~1 € for high reward per correct response). By presenting short sentences, alpha power was measured during listening and maintenance phases of the auditory task to assess listening effort objectively. It was predicted that for the lowest and highest SNRs (-8 and +4 dB), there would be no changes in effort due to reward because the task demand would be too hard and too easy, respectively. However, for the intermediate SNRs (-4 and 0 dB), reward and SNR should interact to modify listening effort with increasing reward leading to increased listening effort. Although a conjecture might be that if more effort is invested due to higher reward then it should also improve the performance, but our expectation based on the results of Koelewijn et al., (2018) was that performance would remain unchanged due to reward.

2.2 Study design

2.2.1 Participants

The participants for this study were 16 (8 females) native Danish-speaking adults with an average age of 25.8 ± 2.8 years who provided written consent prior to study start. Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark. One additional participant was excluded due to non-compliance with experimental instructions. None of the participants suffered from any neurological or hearing disorders. To make sure they were within normal-hearing criteria, the pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz (PTA4) were tested and confirmed to be below 25 SPL HL.

2.2.2 Apparatus

The experiment was set up in a double-walled sound-proof booth. Five loudspeakers were positioned around the participants. The target was presented from a loudspeaker at 0° azimuth in front of the listener. The background noise, consisting of four talkers, was presented from four loudspeakers located at ±90° and ±150° azimuth. All the loudspeakers were 1.2 m away from the listener. The spatial setup of the test is illustrated in **Fig. 2.1**.

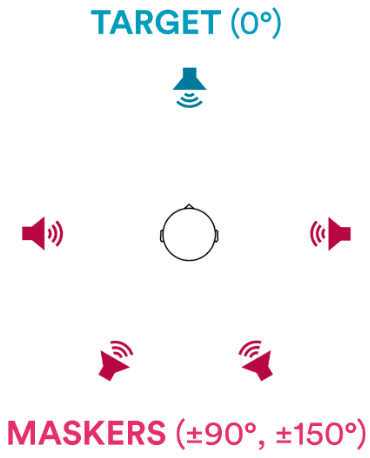


Fig. 2.1 Spatial setup of the task; Target was presented from a loudspeaker at 0° azimuth in front of the listener (in blue) and background noise was presented from four loudspeakers located at ±90° and ±150° (in red). The loudspeakers were 1.2 m away from the listener.

Stimuli were routed through a sound card (RME Hammerfall DSB Multiface II, Audio AG, Haimhausen, Germany) and were played via Genelec 8040A loudspeakers (Genelec Oy, Iisalmi, Finland). EEG data were recorded by a BioSemi ActiveTwo amplifier system (Biosemi, Netherlands) with a standard cap including 64 surface electrodes mounted according to the international 10-20 system with a sampling frequency of 1024 Hz. The cap included DRL and CMS electrodes as references for all other recording electrodes. All electrodes were mounted by applying conductive gel to obtain stable and below 50 mV offset voltage.

2.2.3 Stimuli

Danish Hearing in Noise Test (HINT; Nielsen and Dau, 2011) sentences were used as the target stimuli in the presence of 4-talker babble noise (2 females and 2 males reading Danish text from a newspaper). The position of each babble talker changed randomly from block to block. The A-weighted SPL of the babble was fixed at 70 dB overall (64 dB each). The A-weighted SPL was measured by BK 2250 sound-level meter and BK4231 calibrator. The SPL for each loudspeaker was measured separately and together using unmodulated white noise (spectrally shaped to the stimuli). The level of the target was varied across blocks from 62-74 dB to generate 4 different SNRs: -8, -4, 0 and +4 dB.

2.2.4 Procedure

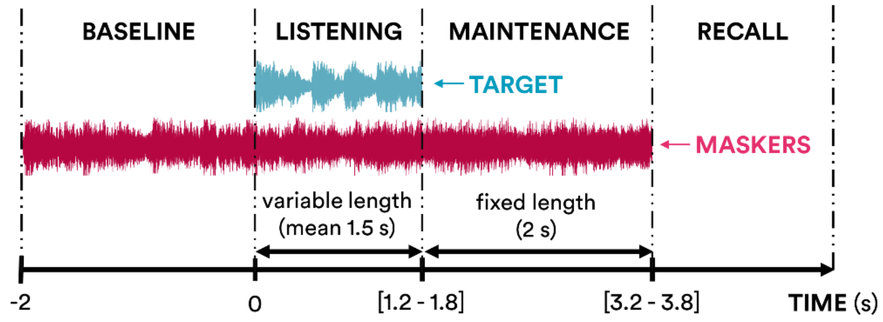
To assess the effects of task demand, 4 different SNRs were chosen (-8 dB, -4 dB, 0 dB and +4 dB) and to assess the effects of motivation, 2 different levels of reward were chosen (low = 0.5 DKK or ~0.06 €, high = 7.5 DKK or ~1 € per correct response). The experiment consisted of 160 trials, divided into 8 separate blocks. Each condition comprised one block of trials (20 trials). The order of conditions was randomized within and between the participants. There was also a training block (20 trials) in the beginning of the experiment (+4 dB SNR with no reward) to familiarize participants with the main experiment.

The participants were informed verbally at the beginning of each block (and also by a written sign visible throughout the block as a reminder) about the reward level of the block (either low or high). Each trial started with 2 s of background noise which was used as the baseline for EEG analysis (baseline phase). After that the HINT sentences were played in the presence of the continuous background noise, during which test subjects were asked to attend to the target (encoding or listening phase). Due to the different lengths of the HINT sentences, this listening phase had a variable duration of between 1.2-1.8 s (mean 1.5 s). After the sentence was finished, the background noise continued for another 2 s during which participants needed to maintain the sentence they just listened to (maintenance phase). When the background noise stopped, the participants were instructed to repeat the sentence (recall phase). If they could repeat all the words within the sentence correctly, then they achieved the reward for that trial (i.e., sentence-based scoring). No feedback was given at the end of trials or blocks, in order to avoid excitement/disappointment (possible confounding motivation factor) from correct/incorrect answers. Only when the experiment was complete were participants informed about the number of sentences they recalled correctly, and hence the money they had earnt overall. The procedure for each trial is illustrated in **Fig. 2.2**.

2.2.5 EEG analysis

There are three major steps in the EEG processing:

1) Preprocessing which includes cleaning the data, resampling and re-referencing
2) Segmentation to extract different phases of the trial
3) Power extraction using wavelet in different frequency bands



4)

5) **Fig. 2.2** Trial procedure: Each trial started with 2 seconds of background noise (baseline). After that the target sentences were played in the presence of background noise (listening) which lasted between 1.2 s to 1.8 s. After the target sentence was finished, the background noise continued for another 2 s (maintenance). When the background noise was stopped, the participants were instructed to repeat the sentence (recall).

2.2.5.1 Pre-processing

Power line noise in the data was rejected with a 50-Hz notch filter with a quality factor of 25. After that, a 3rd-order zero-phase Butterworth bandpass filter with cutoff frequencies of 1-40 Hz was applied to the data and the resulting signals were down-sampled to 256 Hz. Bad channels were detected by visual inspection. On average 6.4 channels were detected as bad channels per participant and they were interpolated using spline interpolation. The eye movements and other sources of unwanted spikes in the data were removed with the joint decorrelation method (De Cheveigné & Parra, 2014). Bad trials were rejected by visual inspection. In total 10.1% of all trials in the study were rejected and no participant had more than 23.7% of trials rejected. Finally, the resulted signals were common re-referenced. The codes used for this part were borrowed from EEGLAB (Delorme & Makeig, 2004), Fieldtrip (Oostenveld et al., 2011), and NoiseTools (de Cheveigné & Arzounian, 2018) toolboxes in addition to custom-written codes.

2.2.5.2 Segmentation

Due to the different lengths of HINT sentences (1.2-1.8 s), the duration of trials across conditions and participants were unequal. This can raise several potential problems for EEG analysis (Luck, 2005), with unequal trial durations introducing unequal variance and bias estimates of power spectral density of the EEG signal. To overcome this issue, the shortest duration of the Danish HINT stimuli was identified (1.2 s) and all the EEG signals during the listening phase were cut to this number (i.e., segments were cut to include only the first 1.2 s of each stimulus regardless of actual duration). Therefore, each trial for EEG analysis consisted of 2 seconds of baseline (-2 s to 0 s), and 1.2 seconds of listening (0 s to 1.2 s). Varying durations were thus removed at the end of the listening period, which also meant that the onset of the maintenance phase varied in its relation to stimulus onset. It should be noted that the maintenance phase, regardless of the length of the target, always lasted for 2 seconds. For this reason, hereafter, the maintenance phase is shown as 1.2 s to 3.2 s in the Results section.

2.2.5.3 Power extraction

Event-related spectral perturbation (ERSP) (Makeig, 1993) was used to evaluate how the EEG power spectra changed over time relative to the baseline. The baseline was chosen as -1.9 s to -0.1 s of the background noise masker, prior to the onset of the target speech. The first 100 ms (i.e., -2 s to -1.9 s) was removed because of the low-level evoked response potential (ERP) due to the start of the sound and the last 100 ms (i.e., -0.1 s to 0 s) was removed because of spectral leakage from the onset of the stimulus. To obtain time–frequency representations of the trials, EEG data were convolved with Morlet wavelets (7 cycles width) in a frequency range between 2 and 35 Hz centered at 100 ms steps within a trial, and then ERSP was calculated using formula (1):

$$ERSP_{t,f}(\%) = \frac{A_{t,f} - R}{R} \times 100 \qquad (1)$$

In which, the ERSP changes are calculated as a percentage. $A_{t,f}$ is the absolute power of the post-stimulus signal in the time window t, and frequency range of f and R is the absolute power of the baseline signal.

EEG power is a 3-dimesion data in space, time and frequency. In order to extract any comparable feature among the manipulations of the study, a specific range for all of the three dimensions needs to be defined.

For the space dimension, based on the initial hypothesis of exploring changes of alpha power in the parietal region, the ERSPs of surrounding electrodes (consisting of CPz, CP1, CP2, CP3, CP4, CP5, CP6, Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO7, PO8) were averaged together to get a more robust estimation of that region. For this particular study, as a strong activity around the frontal lobe was observed, electrodes of that area (AFz, AF3, AF4, Fz, F1, F2, FCz, FC1, FC2) were also averaged together as an additional analysis.

For the frequency dimension, various frequency ranges of power estimation were chosen based on the distinct ERS and ERD in the grand average spectrum (Cohen, 2014). In the listening phase there was one negative peak at 11 Hz which we considered as alpha band (6-13 Hz; **Fig. 2.4**). In the maintenance phase there were three peaks: a positive peak at 7 Hz, a negative peak at 11 Hz, and a positive peak at 14 Hz. Based on this observation we divided our analyses into three different subgroup: low alpha (6-8 Hz; **Fig. 2.5**), high alpha (9-13 Hz; **Fig. 2.6**) and beta (14-18 Hz; **Fig. 2.7**).

For the time dimension, the ERSP time window in the listening phase was chosen from 0.1 s to 1.1 s, and in the maintenance phase from 1.3 s to 2.8 s. The first 100 ms in both phases were cut to avoid getting parts of low-level ERPs. The last 100 ms in the listening phase (i.e., 1.1 s to 1.2 s) was removed because of spectral leakage from the onset of the maintenance phase. The last 400 ms in the maintenance phase (i.e., 2.8 s to 3.2 s) was removed due to time-frequency trade off in the wavelet analysis where wavelet cycles in lowest frequencies could not contain the signal around the edges. Due to participants' different reaction times in responding, EEG in the recall phase could not be analyzed.

2.2.6 Self-report

At the end of the testing, each participant responded to a declarative sentence (presented in English), *"My motivation to do this task was affected by the scale of the reward"* on a five-point Likert scale ("Disagree", "Slightly disagree", "Neutral", "Slightly agree", "Agree"). The aim of this rating was to see how much each participant perceived themselves to be motivated by the change in monetary reward in the experiment.

2.2.7 Statistics

Linear mixed model (LMM) analyses were used for statistical evaluation of the results. The advantage of LMM is that it can model fixed effects, as well as random effects within the experiment which provides more flexible functionality for model estimation (DeBruine & Barr, 2021). In this study, LMM was implemented to analyze task performance and EEG power in different bands and phases. In all the models, SNR and reward were treated as independent measures, thus as fixed factors, with participants as random factors. SNR values were centred around 0 (i.e., -6, -2, +2, +6 dB) solely for the purposes of the LMM in order to avoid correlation between the main effects and interactions in the model (Harrison et al., 2018). The estimates of the LMM (β), standard error (SE), t-values and degrees of freedom (t_{DF}) and P values are reported in the Results section.

For the performance in the task (i.e., sentence recognition), the main effects of SNR and reward and their linear interaction ($Performance \sim 1 + SNR \times Reward + (1|Subject\ ID)$) are reported in the Results section. For the EEG power, an additional quadratic term ($Power \sim 1 + SNR^2 \times Reward + (1|Subject\ ID)$) was added to the LMM. Based on the hypothesis of an interaction between task demand and motivation, a quadratic interaction between SNR and reward is expected, i.e., more pronounced difference in effort in the middle conditions (-4 dB and 0 dB SNR) compared to the extreme conditions (-8 dB and +4 dB SNR). Furthermore, the two models for EEG power ($SNR \times Reward$ vs. $SNR^2 \times Reward$) were compared against each other, using the MATLAB function *compare* to ensure a better fit of the LMM with the quadratic term. For this comparison, likelihood-ratio (Yuanjia Wang & Chen, 2012) with the difference of degrees of freedom (LR_{DF}) and p value are reported in the Results section.

In order to investigate if there is any correlation between performance and EEG power, code scripts from the Robust Correlation Toolbox were borrowed (Pernet et al., 2013). For this purpose, Pearson skipped correlation was used to avoid outliers skewing the results. Also, a bootstrap resampling process was done with 1000 repetitions to mitigate the correlation bias of inter-dependency of samples within an individual (Pernet et al., 2013). To determine whether Pearson coefficient *r* is significant or not, a 95% confidence interval (CI) of bootstrapped data should not contain zero (i.e., the difference of correlation coefficients for the real data compared to shuffled data should be anything but zero). All the statistical tests were performed using the Statistics toolbox of MATLAB 2018b software.

2.3 Results

2.3.1 Performance

The performance accuracy was measured based on sentence scoring meaning that the reward was only given if all the words within the sentence were repeated correctly. The results for performance (**Fig. 2.3**) showed a significant main effect of SNR (β = 6.96, SE = 0.26, t_{124} = 25.90, $p < 0.001$), but no main effect of reward (β = 0.15, SE = 2.40, t_{124} = 0.06, p = 0.94) nor any interaction between SNR and reward (β = -0.21, SE = 0.53, t_{124} = -0.40, p = 0.68).

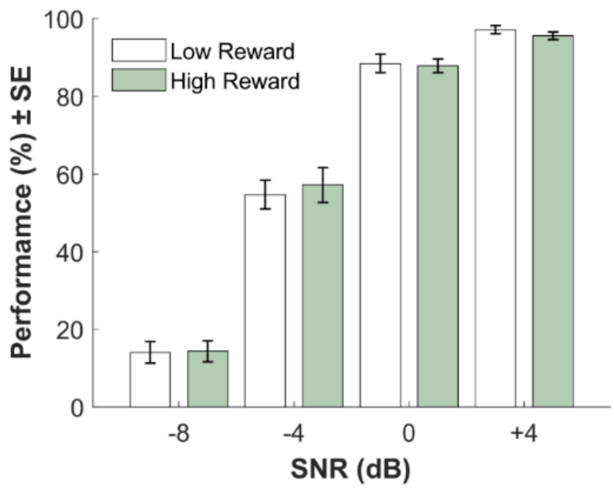


Fig. 2.3 Results of performance accuracy: The percentage of correctly repeated sentences. The error bars show standard error of the mean. There was a significant effect of SNR without any effects of reward.

2.3.2 EEG

2.3.2.1 Listening

In the listening phase there was a desynchronization in alpha power (**Fig. 2.4**), however it did not show any significant changes by SNR and reward in either parietal (**Table 2.1)** or frontal regions (**Table 2.2)**.

2.3.2.2 Maintenance

In the parietal region, there were synchronization of low alpha (**Fig. 2.5)** and beta bands (**Fig. 2.7**), and desynchronization of high alpha band (**Fig. 2.6**) in the maintenance phase. In the low alpha power during the maintenance phase, there was a significant inverted U-shaped pattern of SNR, as well as quadratic interaction between SNR and reward. Additionally, we tested the appropriateness of the model with the quadratic component included vs excluded, finding a nominal improvement in the fit of the LMM when the quadratic was included ($LR_2 = 14.64$, $p < 0.001$). In other words, it can be concluded that 1) effort increased up until the task became too hard, and also 2) when reward was higher, listeners increased their effort in intermediate conditions. In the frontal region, there was also a quadratic effect of SNR in low alpha, but without quadratic interaction between SNR and reward.

Table 2.1 Results of mixed model based on SNR and reward predictors: estimates of relative power changes in the parietal region in different bands and phases. Significant p-values are shown in black.

DF = 122	Listening	Maintenance		
Band Predictor	Alpha	Low Alpha	High Alpha	Beta
SNR	β = 0.42 SE = 0.25 t = 1.65 p = 0.099	β = 0.28 SE = 0.22 t = 1.26 p = 0.207	β = 0.23 SE = 0.27 t = 0.87 p = 0.380	β = -0.33 SE = 0.21 t = -1.52 p = 0.129
SNR^2	β = -0.03 SE = 0.07 t = -0.46 p = 0.640	**β = -0.19** **SE = 0.06** **t = -3.16** **p < 0.001**	β = -0.11 SE = 0.07 t = -1.50 p = 0.134	β = -0.09 SE = 0.06 t = -1.55 p = 0.122
Reward	β = 5.33 SE = 3.64 t = 1.46 p = 0.145	β = 5.88 SE = 3.21 t = 1.82 p = 0.069	β = 5.03 SE = 3.87 t = 1.30 p = 0.196	β = 4.00 SE = 3.11 t = 1.28 p = 0.200
SNR:Reward	β = 0.11 SE = 0.50 t = 0.21 p = 0.827	β = 0.01 SE = 0.44 t = 0.03 p = 0.971	β = 0.23 SE = 0.54 t = 0.43 p = 0.663	β = -0.09 SE = 0.43 t = -0.21 p = 0.827
SNR^2:Reward	β = -0.08 SE = 0.14 t = -0.59 p = 0.553	**β = -0.29** **SE = 0.12** **t = -2.37** **p = 0.018**	β = -0.26 SE = 0.15 t = -1.72 p = 0.086	β = -0.17 SE = 0.12 t = -1.40 p = 0.161

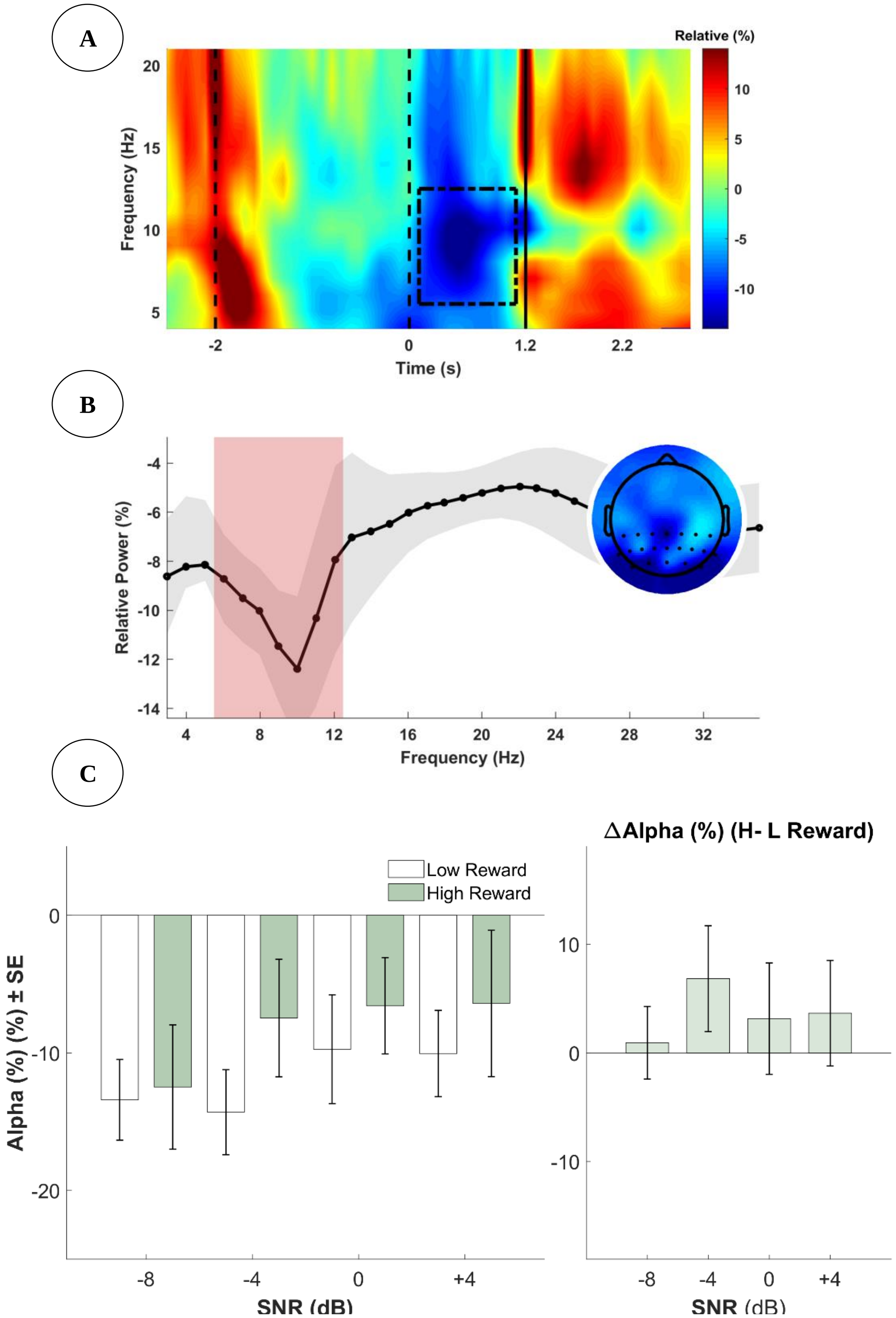


Fig. 2.4 Power changes during listening in the parietal region: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of alpha power by SNR and reward in the highlighted window of panel A which showed no significant modulation. The error bars show standard error of the mean.

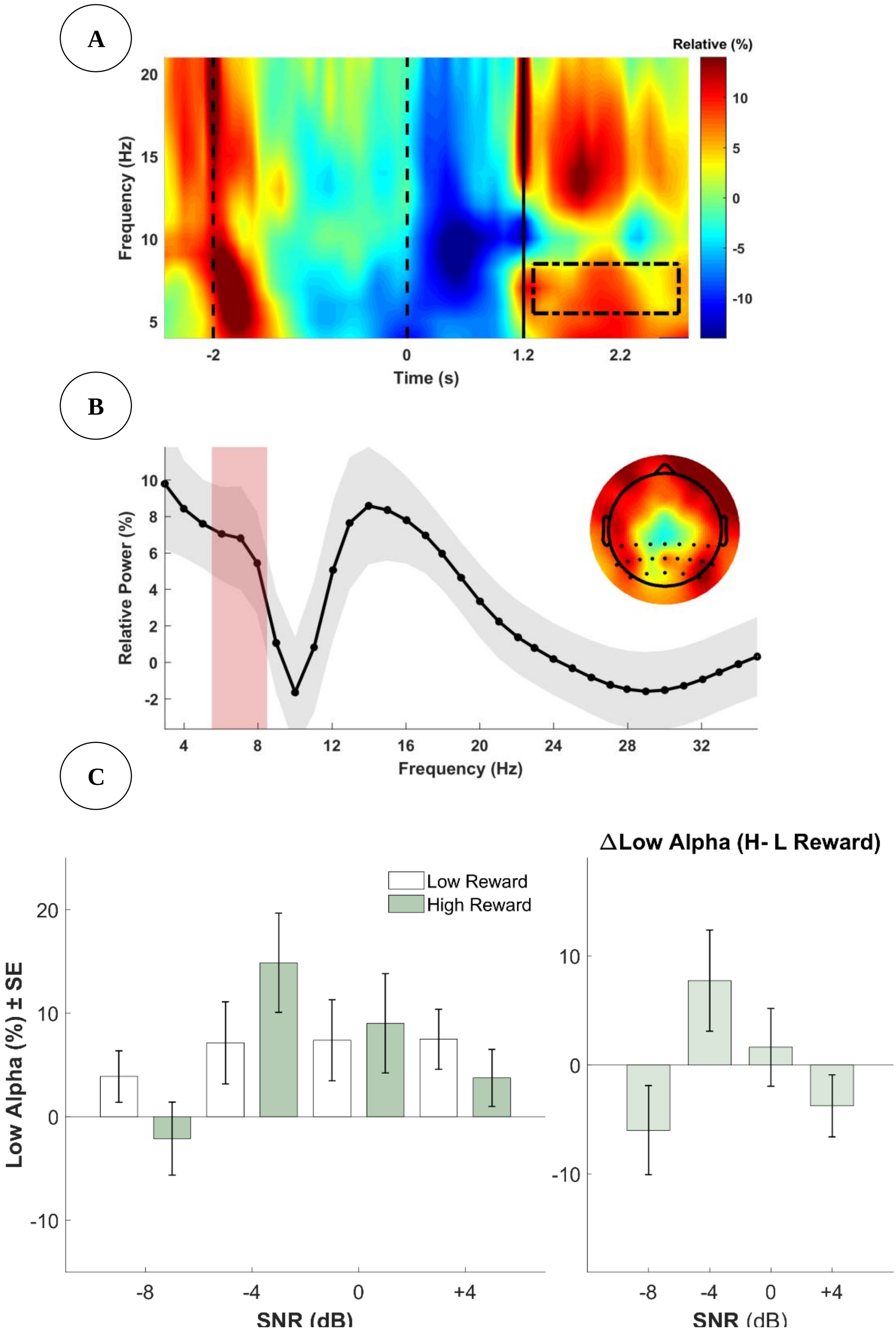


Fig. 2.5 Power changes during maintenance in the parietal region: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of low alpha power by SNR and reward in the highlighted window of panel A which showed significant quadratic interaction between the two. The error bars show standard error of the mean.

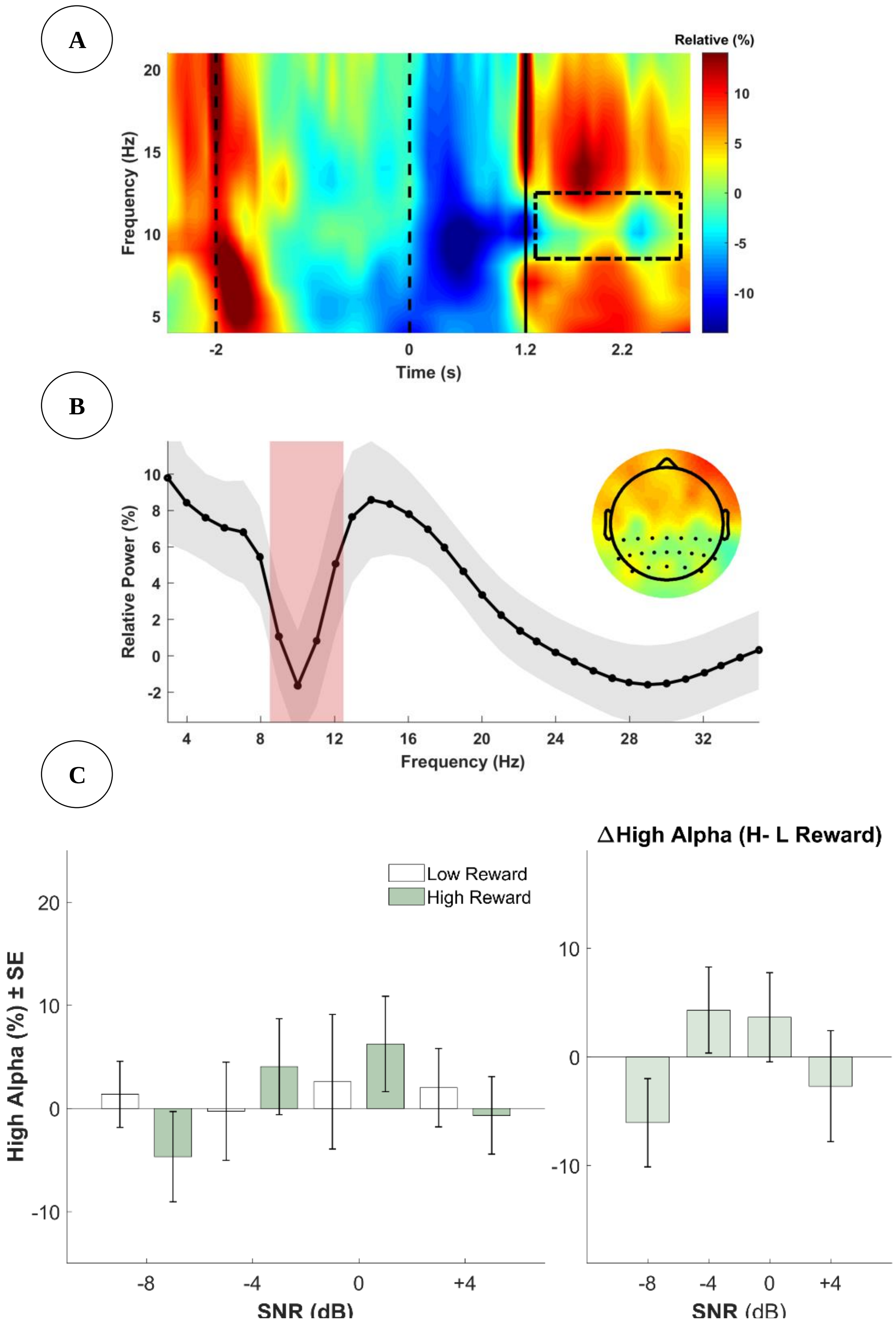


Fig. 2.6 Power changes during maintenance in the parietal region: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of high alpha power by SNR and reward in the highlighted window of panel A which showed no significant modulation. The error bars show standard error of the mean.

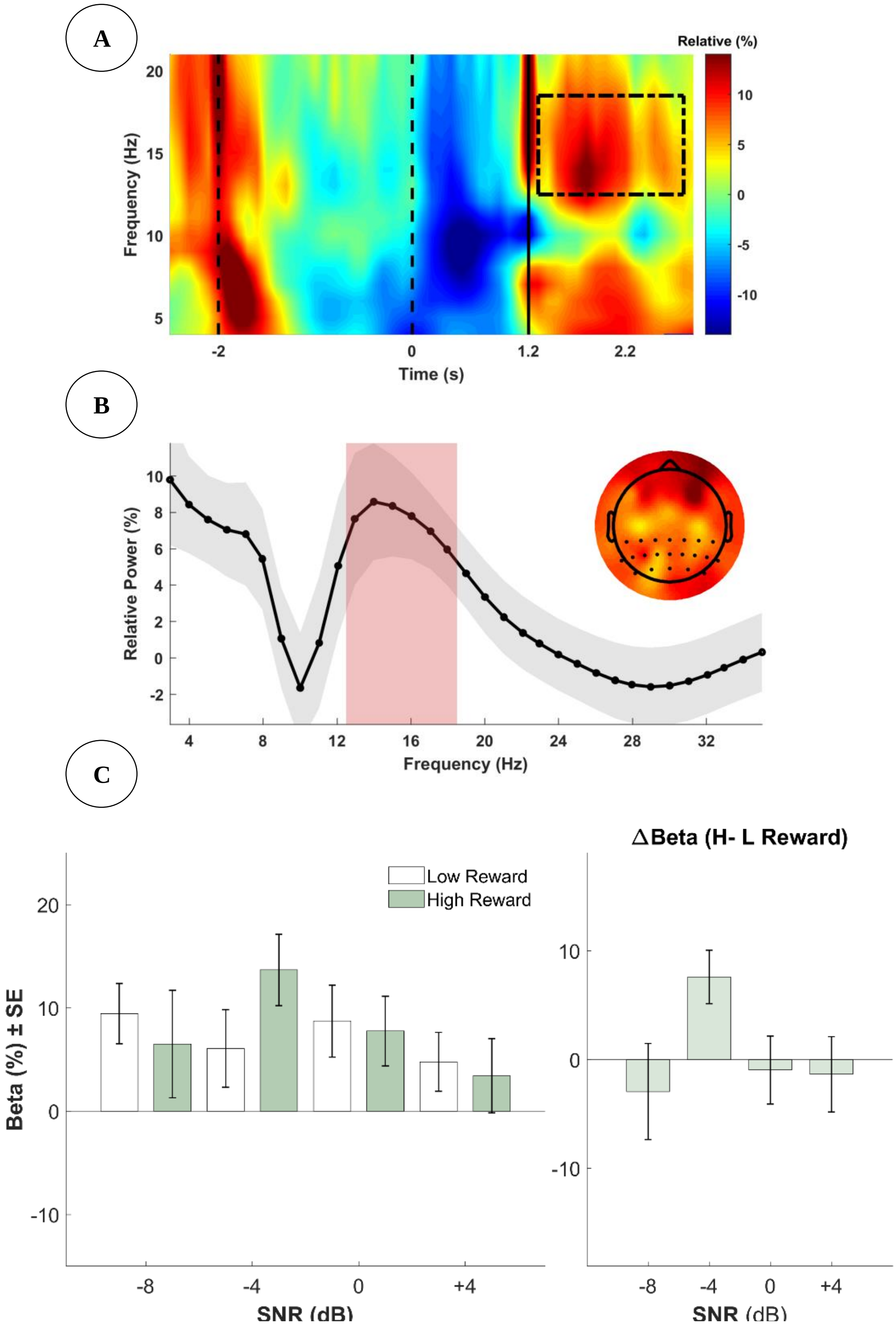


Fig. 2.7 Power changes during maintenance in the parietal region: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR and reward in the highlighted window of panel A which showed no significant modulation. The error bars show standard error of the mean.

Table 2.2 Results of mixed model based on SNR and reward predictors: estimates of relative power changes in the frontal region in different bands and phases. Significant p-values are shown in black.

DF = 122	Listening	Maintenance		
Band / Predictor	Alpha	Low Alpha	High Alpha	Beta
SNR	β = 0.04 SE = 0.29 t = 0.16 p = 0.869	β = 0.11 SE = 0.30 t = 0.35 p = 0.720	β = 0.13 SE = 0.39 t = 0.35 p = 0.724	β = -0.56 SE = 0.34 t = -1.63 p = 0.105
SNR^2	β = -0.08 SE = 0.08 t = -1.04 p = 0.297	**β = -0.18** **SE = 0.08** **t = -2.18** **p = 0.030**	β = -0.17 SE = 0.10 t = -1.61 p = 0.109	β = -0.04 SE = 0.09 t = -0.49 p = 0.623
Reward	β = -0.06 SE = 4.16 t = -0.01 p = 0.987	β = 7.93 SE = 4.43 t = 1.78 p = 0.076	β = 8.14 SE = 5.61 t = 1.45 p = 0.149	β = 9.56 SE = 4.95 t = 1.93 p = 0.055
SNR:Reward	β = 0.19 SE = 0.58 t = 0.33 p = 0.741	β = 0.61 SE = 0.61 t = 0.99 p = 0.320	β = 0.47 SE = 0.78 t = 0.61 p = 0.542	β = 0.97 SE = 0.69 t = 1.40 p = 0.162
SNR^2:Reward	β = -0.02 SE = 016 t = -0.14 p = 0.881	β = -0.33 SE = 0.17 t = -1.94 p = 0.053	β = -0.41 SE = 0.21 t = -1.91 p = 0.058	β = -0.36 SE = 0.19 t = -1.89 p = 0.060

2.3.3 "Motivated" subgroup

When participants were asked whether their motivation was affected by the scale of monetary reward (i.e., the question at the end of the experiment), 1 disagreed, 2 slightly disagreed, 4 responded neutrally, 6 slightly agreed and 2 agreed.

As a post-hoc analysis the minority of participants (3 out of 16) who disagreed/slightly disagreed to being motivated by reward was excluded to check if the alpha power results would be influenced by their absence. To put it briefly, there was no change of significant results as reported by 16 participants (as in **Table 2.1**). Parietal low alpha power still showed significant inverted U-shape pattern due to SNR and significant quadratic interaction between SNR and reward in maintenance. The details of the LMM results for this subgroup is in **Table 2.3**.

Table 2.3 Results of mixed model based on SNR and reward predictors for the "motivated" subgroup: estimates of relative power changes in the parietal region in different bands and phases. Significant p-values are shown in black.

DF = 98	Listening	Maintenance		
Band / **Predictor**	**Alpha**	**Low Alpha**	**High Alpha**	**Beta**
SNR	β = 0.28 SE = 0.28 t = 1.01 p = 0.313	β = 0.28 SE = 0.25 t = 1.10 p = 0.270	β = 0.18 SE = 0.31 t = 0.58 p = 0.557	β = -0.37 SE = 0.24 t = -1.52 p = 0.131
SNR^2	β = -0.11 SE = 0.07 t = -1.49 p = 0.137	**β = -0.21** **SE = 0.07** **t = -3.06** **p = 0.002**	β = -0.17 SE = 0.08 t = -1.93 p = 0.055	β = -0.10 SE = 0.06 t = -1.52 p = 0.130
Reward	β = 5.25 SE = 4.03 t = 1.30 p = 0.195	β = 6.90 SE = 3.62 t = 1.90 p = 0.059	β = 5.78 SE = 4.51 t = 1.28 p = 0.203	β = 4.49 SE = 3.55 t = 1.26 p = 0.208
SNR:Reward	β = 0.05 SE = 0.56 t = 0.09 p = 0.927	β = -0.08 SE = 0.50 t = -0.17 p = 0.865	β = -0.02 SE = 0.63 t = -0.03 p = 0.972	β = -0.01 SE = 0.49 t = -0.01 p = 0.998
SNR^2:Reward	β = -0.03 SE = 0.15 t = -0.21 p = 0.832	**β = -0.31** **SE = 0.14** **t = -2.25** **p = 0.026**	β = -0.24 SE = 0.17 t = -1.38 p = 0.170	β = -0.21 SE = 0.13 t = -1.54 p = 0.126

2.3.4 Correlations

The analysis of Pearson skipped correlation showed that there was no significant correlation between performance and the power of different bands in any phase (**Table 2.4**). The correlation graphs with their respective *r* and 95% CI and bootstrapped resampling are shown in **Fig. 2.8**.

Table 2.4 Pearson skipped correlation between performance and EEG power in the parietal region in different bands and phases.

	Listening	Maintenance		
Band / **Electrodes**	**Alpha**	**Low Alpha**	**High Alpha**	**Beta**
Parietal	r = 0.07 CI = [-0.14 0.25]	r = 0.17 CI = [-0.01 0.33]	r = 0.08 CI = [-0.10 0.25]	r = -0.01 CI = [-0.18 0.14]

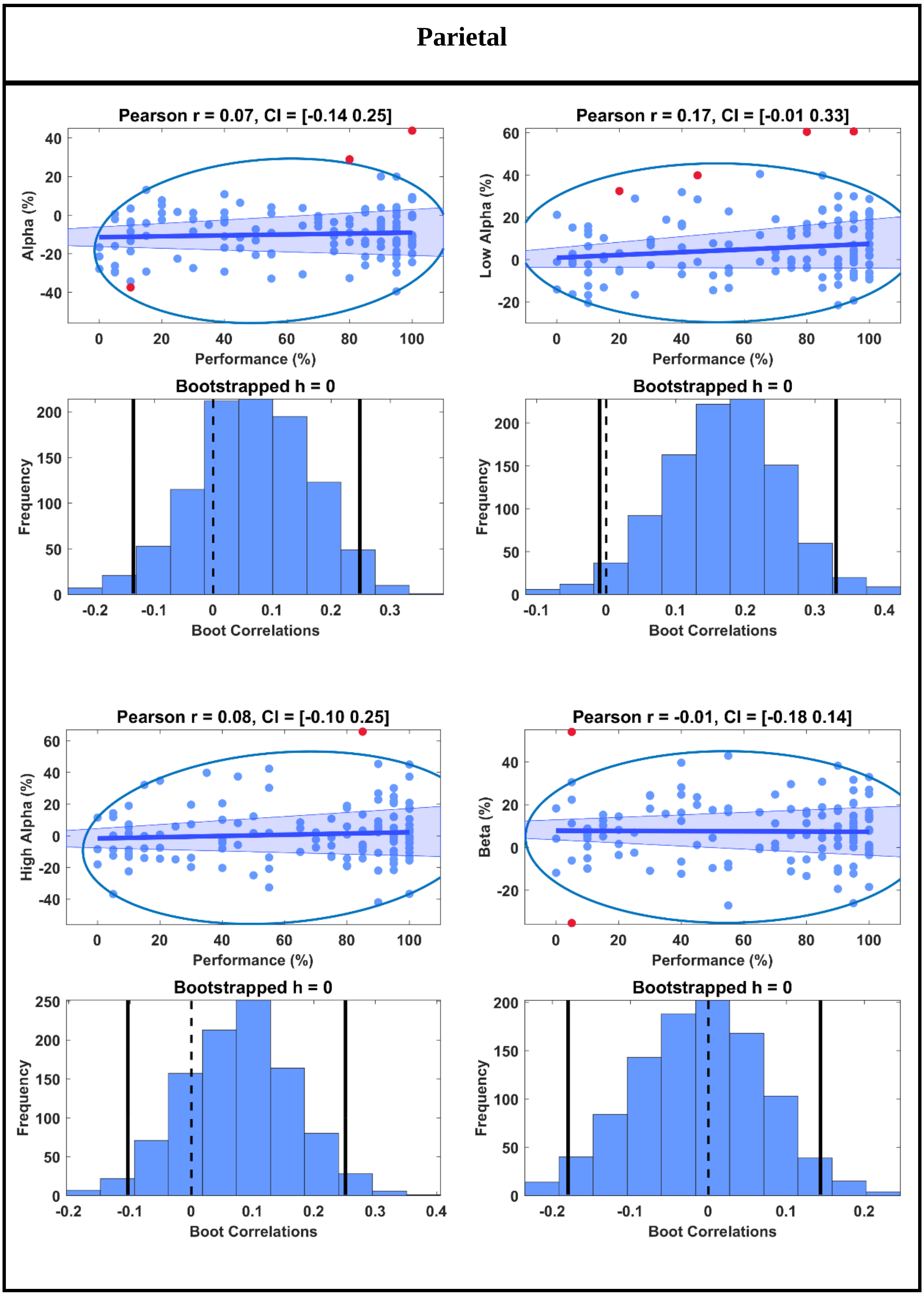


Fig. 2.8 Pearson's skipped correlation between performance and EEG power during maintenance (low alpha, high alpha and beta) and during listening (alpha). The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

2.4 Discussion

2.4.1 Overview

The results of this study showed that in in extreme SNRs (-8 and +4 dB) higher reward may not lead to increased effort. However, it was in the middle SNRs (0 and -4 dB) that increased parietal low alpha was observed during maintenance phase which suggested participants may have spent more effort in those conditions. Regardless of increased effort, there was no behavioural benefits due to reward.

2.4.2 No behavioural benefit of reward

It could be expected that when people are more motivated to do a task (such as through greater reward), they spend more effort in the hope of achieving a better outcome (Brehm & Self, 1989). However, in the current study higher reward did not lead to better performance even in the SNRs where increased effort was observed. While previous studies (Koelewijn et al., 2018) and theories (Lunner et al., 2016) speculate that listening effort may vary even if the performance is unchanged, it is also possible that our performance measure was insufficiently sensitive. Similar to Koelewijn et al., (2018), performance was measured based on whole sentences and not individual words, which may have missed word-level differences in performance due to reward. Another explanation could be that having block-based or trial-based reward could make a difference in overall performance. Trial-based cues have shown to help transient increases in preparatory effort which resulted in improving behavioural performance compared to sustained motivation (Kostandyan et al., 2019). Hence the block-based format of our manipulation may have led to a conservative sustained effort throughout the block. While this limitation could be addressed in future, our finding of changes in alpha power (interpreted as effort) in spite of no changes in performance nonetheless points to the importance of using objective measures of listening effort in addition to measures of performance.

2.4.3 Alpha during maintenance

The analyses of EEG data showed that varying SNR and level of reward during a listening task led to changes in low alpha power primarily in the maintenance phase, and not in the listening phase. The importance of the alpha activation in the maintenance phase (i.e., after speech signal offset) has been shown in other studies such as Obleser et al., (2012), which reported alpha oscillations to be affected by both speech degradation and working memory load. This could be due to difficult listening situations resulting in noisy memory representations and word mismatch in the working memory, which puts more strain on the STM while the sentence is being held in mind (Wilsch & Obleser, 2016).

Given that the maintenance phase occurred in the presence of background noise, the activation of the low alpha band could very well serve a function of inhibiting unattended sound sources (Wilsch & Obleser, 2016) to facilitate the retention of sentences. Lower SNRs cause more word mismatches during listening, which draws more resources from working memory to identify them. However, when the task became too difficult (-8 dB), the participants started to disengage and low alpha activity started to decrease. Such an inverted U-shaped curve has been shown in numerous effortful listening studies, using both EEG (Decruy et al., 2020) and pupillometry (Ohlenforst et al., 2018; Wendt et al., 2018).

2.4.4 Alpha during listening

In contrast to the maintenance phase, alpha activity during the listening phase did not change with manipulation of SNR and reward. One reason for this observation might be that EEG power in this phase (which lasted around 1 s) reflects task engagement rather than stimulus features. Despite lack of statistical significance, it is important to note that unlike maintenance phase, there is a wide frequency range of desynchrony during listening which reflects neural excitability that emerges from active processing of information (Klimesch, 2012; Klimesch et al., 2007; Palva & Palva, 2007; Wilsch & Obleser, 2016). It has also been suggested that such wide desynchronization represents maximal information capacity (Pfurtscheller, 2001).

During the maintenance phase of a listening task, it has been shown that alpha power changes are sensitive to subtle manipulations of memory load, such as monotonic increase in spectral amplitude (Leiberg et al., 2006) or in a delayed-match-to-sample paradigm (Luo et al., 2005). However, these studies have not reported equivalent effects during their short listening phases (~1 s), which may be another indication that alpha power during early listening phase is more of a global feature and shows task engagement. In fact, studies that appear to show alpha power changes according to task demand during the listening phase are those with a longer duration of target speech such as Wöstmann et al., (2015) (~3 s) or Seifi Ala et al., (2020) (~30 s). It is possible that longer auditory stimuli might be required to look for alpha changes to track listening effort. This is explored further in **Chapter 4** and **Chapter 5**.

2.4.5 Interaction of reward and SNR

Previous listening studies have shown different effects of monetary reward on listening effort (e.g., Koelewijn et al., 2018; Picou & Ricketts, 2014). Similar to the current study, Koeleweijn et al. (2018) investigated the effects of motivation and task demand on listening effort in a speech-in-noise task, although effort was measured using pupillometry. They also used two levels of monetary reward (high vs low), but only two levels of task demand (50% and 85%-correct thresholds in an adaptive sentence-recognition procedure). They observed higher reward increased effort as measured by peak pupil dilation, but there was no (linear) interaction between reward and task demand. To compare the current results with Koelewijn et al., I analyzed the effects of reward and SNR for only the two SNR conditions of our study (-4 and 0 dB) which were very similar in speech intelligibility (56% and 88%, respectively) to Koelewijn et al. (2018). Using LM, I did not find any effects of reward ($\beta = 4.68$, SE = 2.89, $t_{60} = 1.62$, $p = 0.110$) or interaction between reward and SNR ($\beta = -1.52$, SE = 1.44, $t_{60} = -1.05$, $p = 0.295$). Using the same statistical approach as Koelewijn et al. (a repeated measure ANOVA), I could not find any significant effects of reward ($F(1,15) = 1.92$, $p = 0.185$) nor interaction between reward and SNR ($F(1,15) = -1.61$, $p = 0.223$). Therefore, to observe the interactive effects of reward and task demand on effort, as hypothesized in FUEL, it may be necessary to measure at a range of task demands spanning relatively difficult, intermediate and easy levels. However, in the study by Plain et al., (2020), even with a wide range of task demand, no interactive effect of reward on effort was observed in PEP data. Despite similar reward values to Koelewijn et al., (2018), they suspected that longer sessions in their design required higher monetary reward to increase effort compared to short sessions in Koelewijn et al., (2018).

2.4.6 Limitations

One of the reasons that there were no behavioural benefits of reward in this experiment, might have been due to the fact that evaluation of performance was sentence based. That means a response with 4 out of 5 correct words was considered the same as a response with no correct words (both considered as wrong answers). Scoring the performance based on words might have been a more accurate approach to evaluate behavioural benefits. Future studies are needed to see how the difference between word and sentence scoring would influence any conclusions on the behavioural benefits of monetary reward.

The other shortcoming was using a general questionnaire at the end of the experiment instead of a condition-specific questionnaire. While the participants declared whether they were motivated by scale of reward or not, they were not specifically asked if that varied across conditions. For this reason, this questionnaire could not be compared to other measurements of the study beyond what was done in **Section 2.3.3** (by removing the "unmotivated" participants).

2.5 Conclusion

The hypothesis of this study was that higher reward would lead to increased listening effort when the task demand is moderate (i.e., not too high – leading to floor effects – or too low – leading to ceiling effects). To test this hypothesis, four different SNRs (-8 dB, -4 dB, 0 dB and +4 dB) with two levels of reward (high and low) were presented to participants using Danish HINT speech material and EEG signals were recorded during the task. The normalized posterior low alpha (6-8 Hz) during the maintenance showed a quadratic interaction between SNR and monetary reward. In the intermediate (-4 and 0 dB SNR) conditions, increased effort was observed in the high reward condition compared to the low reward condition. However, in the very easy (+4 dB SNR) and very difficult (-8 dB SNR) conditions there was no such effect. These results show that both environmental factors (task demand) and personal factors (motivation) are important for the evaluation of listening effort.

Chapter 3
Listening effort in simulated rooms

The current study was a joint project with the fellow PhD student Sergio Aguirre at University of Nottingham, Hearing Sciences – Scottish Section. The contributions were as followed:

TSA: Study design, preparing set up, data collection, data analysis, data interpretation.

SA: Study design, room simulations, sound calibration, preparing set up, data collection, data interpretation.

3.1 Introduction

In the previous chapter I talked about how motivation (a personal factor) can interact with SNR (an environmental factor) to change listening effort. In this chapter, I will focus on interaction of SNR and rooms (two environmental factors) to investigate how they shape listening effort.

3.1.1 Reverberation

Sound generated by a source in an enclosed space can reach listener either directly (free field) or indirectly (reverberant). Reverberant fields are generated by reflected sounds from enclosed spaces and are time dependent. When the source suddenly ceases, a sound field persists for a finite interval as the result of multiple reflections and the low velocity of sound propagation. This residual acoustic energy constitutes the reverberant field. Reverberation time (RT) is a characterization of the acoustics of a space that represents the amount of time required to dissipate the energy of a sound source by 60 dB (T_{60}) after the sound source has ceased (Rossing, 2007). The remaining sound energy distorts the envelope and fine structure of the received sound (Ratnam et al., 2003) and can blur auditory cues, rapid transitions between phonemes, decrease low frequency modulation of a signal, and may compromise speech intelligibility (Hazrati & Loizou, 2012). Moreover, the reverberation time is frequency dependent. How walls/objects absorb lower (audible) frequencies is often very different from higher frequencies. This not only affects the T_{60} across frequencies, but how the reverberation affects speech intelligibility.

Interest in studies with more ecologically valid test situations has encouraged researchers to include reverberation in hearing tests, as many listening situations in closed spaces are reverberant. Anechoic chambers that are often used in auditory studies have a T_{60} of close to 0. More realistic spaces (such as rooms/offices with furniture and people) have T_{60} of less than 1 s at 1 kHz (Knecht et al., 2002). Churches, cathedrals, and other large spaces with hard surfaces have T_{60} of larger than 2 s (Desarnaulds et al., 2002). Therefore, simulated rooms in anechoic chambers can provide the opportunity to study reverberation in more realistic enclosed spaces.

In reverberant environments, speech understanding is difficult, especially for hearing-impaired individuals. A study by Xia et al., (2018), evaluated the effects of noise and reverberation on speech intelligibility in normal hearing and hearing-impaired individuals. Sixteen acoustic scenes with four different reverberant rooms (T_{60} = 0, 0.66, 0.8, 1.88 s) and four acoustic backgrounds (quiet, SNR = 5, 10 dB, one-talker speaker) were simulated. The initial results showed that speech intelligibility for hearing-impaired listeners was less than normal hearing listeners. However, when the ceiling effect was corrected for, the difference between the two groups became much smaller. Based on these results, the authors suggested that part of the difference in susceptibility to reverberation between normal-hearing and hearing-impaired listeners could be due to ceiling effects.

There are also studies that have investigated the effects of reverberation on listening effort subjectively and/or objectively. In a study by Prodi & Visentin (2019), participants performed speech-in-noise task in three different reverberant conditions (T_{30} = 0, 0.3, 0.65 s) with stationary or fluctuating noise and SNR varying from –14.9 dB to 4 dB. The task was to select the presented word among rhyming words which appeared on the screen. Listening effort was measured objectively with response time. The results showed significant increase of response time due to higher reverberation and noise level which showed increased effort for these conditions. The authors also found an interaction between reverberation and noise levels which suggests reverberation can affect listening effort differently depending on the noise type and level.

Importantly, reverberation can affect listening differently depending on age. In a study by Kwak et al., (2018), young and elderly adults participated in a sentence recognition task with four different SNRs (0, +3, +6 dB and quite) and five levels of RT (0, 0.5, 1, 1.5, 2 s). Intelligibility was measured via a sentence recognition task, and listening effort was evaluated subjectively via a questionnaire. The results showed that decreasing SNR and increasing RT led to poorer sentence recognition and increased listening effort in both groups. However, RT affected the sentence recognition performance and listening effort more in the elderly group than in the younger group. This study suggests that listening in reverberant conditions may get worse with age. Opposite to the studies by Prodi & Visentin (2019) and Kwak et al., (2018), there are studies that have failed to find evidence that reverberation affects listening effort. For example, in a study by Picou et al., (2016), response time in a dual-task paradigm was used as a behavioural measure of listening effort. Participants were tested with variant SNRs in three levels of reverberation (T_{30} < 0.1, 0.4, 0.8 s). While word recognition was negatively affected by SNR and reverberation, response times only showed a significant effect of SNR. Therefore, reverberation in this study did not reveal any effect on listening effort. While inconsistent with previous results, the authors speculated that perhaps their young normal hearing participants were not sensitive enough to the difference of applied reverberant conditions in this study.

3.1.2 Subjective vs. objective measures of effort

In **Section 1.3.1**, it was mentioned that subjective and objective measures of listening effort may or may not agree with each other during an auditory task; similar conflicting findings have been found in reverberation studies. For example, in the study by Holube et al., (2016), reverberation effects on listening effort were found to be different using questionnaire (subjectively) and electrodermal activity (objectively). In this study, young normal-hearing and elderly hearing-impaired participants listened to sentences either with stationary background noise or with reverberation. Two levels of SNRs were used for creating easy listening situation (6 dB for normal-hearing and 10 dB SNR for hearing impaired) and difficult listening situation (-6 dB for normal-hearing and -2 dB SNR for hearing impaired). Similarly, two levels of RT were used for creating easy listening situation (0.5 s for normal-hearing and 0.3 s for hearing impaired) and difficult listening situation (4 s for normal-hearing and 2 s for hearing impaired). Easy situations led to 100% speech intelligibility whereas difficult listening situations led to 30%-80% speech intelligibility. While subjective ratings showed significant differences between the easy and hard conditions in both groups, the electrodermal activity revealed no significant trends. It is possible that electrodermal activity should be regarded as a measure of autonomic stress reaction that could not capture listening effort in this context.

However, there are studies that show subjective and objective measures of listening effort align with each other in reverberant conditions. In the study by Picou & Ricketts, (2018) adults with symmetrical sensorineural hearing loss participated in a dual-task paradigm with three microphone settings in a reverberant condition (T_{30} = 0.7 s). Using questionnaire (subjectively) and response time (objectively), the authors found that subjective rating aligned with response time as a measure of listening effort. The authors suggested that their conflicting results with their previous study where reverberation did not affect listening effort (Picou et al., 2016) was perhaps due to small RT coupled with the use of monosyllabic words in the earlier study that did not lead to any effects of reverberation on listening effort. However, it is notable that this study only used one reverberation time and did not elicit a large range in intelligibility performance, indicating that task demands were fairly similar throughout the study.

The reason that subjective and objective measures of listening effort show conflicting results across studies might be because they capture different processing in the brain. I speculated in **Chapter 1** that using a wide range of task demand might be the key in revealing the differences between subjective and objective ratings of listening effort. Subjective measures mostly change linearly with task demand, but if the task demand is large enough, an inverted U-shaped pattern of objective effort is expected. One plausible explanation is that subjective ratings of effort may be more related to speech intelligibility rather than expenditure of resources in the brain. In this chapter, the idea that subjective and objective measures of listening effort differ from each other when the range of task demand is large enough is put into test.

3.1.3 Objectives

The objective of this chapter is to investigate how SNR and reverberation interact with each other within a simulated room to affect speech intelligibility and listening effort. The hypothesis is that increasing RT would increase the SNR point where the listeners disengage from the task (i.e., higher RT leads to a faster disengagement due to added task demand). For this purpose, three different SNRs (-8, -3, +2 dB) and three simulated rooms (RT = 0, 0.5, 1.1 s) were used. Speech intelligibility was measured based on word scoring and listening effort was measured both subjectively and objectively. For the subjective measure of effort, a questionnaire was used. For the objective measure of effort, low alpha power in maintenance was used based on the findings in **Chapter 2**. However, due to our focus on identifying appropriate measures of objective listening effort, I additionally explored the listening phase, and other power bands that seemed relevant in the maintenance phase. I expected to observe speech intelligibility and subjective ratings linearly change with SNR and RT. However, for low alpha power in the maintenance, I expected to see an inverted U-shaped pattern because of the disengagement that may occur in the most demanding conditions in the study.

3.2 Study design

3.2.1 Participants

18 normal-hearing native Danish-speaking adults (8 females) with an average age of 36.9 ± 11.2 years participated in this study. All the participants signed a written consent form prior to the test. One participant was positioned incorrectly in the sound field, so his data was discarded and the data for the other 17 participants were used for further analysis. Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark. For each participant, the pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz (PTA4) were tested and confirmed to be below 25 SPL HL.

3.2.2 Apparatus

The experiment was set up in an anechoic room with inner dimensions of 4.3 x 3.4 x 2.7 m. The experimental setup consisted of a circular array of 24 loudspeakers positioned on 15° azimuth intervals with 1.2 m distance to the centre of the circle where the participant was seated. The height of the seat was adjusted so that the loudspeaker drivers were at ear level. The target and maskers were simulated to be at the same distance as the actual loudspeaker array, with the target at 0° and maskers at 90°, 150°, 210° and 270°. The position of the participant during all testing was monitored using a laser pointer and a camera to ensure they are placed in the middle of loudspeakers throughout the test.

Stimuli were routed through a sound card (MOTU PCIe-424) with an Firewire 440 connection to the MOTU Audio 24I/O interface and were played via 24 calibrated loudspeakers (Genelec 8030A/C) described above. The EEG apparatus was as described in **Section 2.2.2**.

3.2.3 Stimuli

The speech materials, including target and 4-talker background noise, were from the Danish HINT, as described in **Section 2.2.3**. The overall maskers' SPL was set at 70 dB (64 dB for each background talker), and the targets were set at 62 dB, 67 dB and 72 dB to generate three different SNR conditions: -8, -3, and +2 dB. In this study, SNR was defined as the input SNR, which means the long-term average sound level of the target signal compared to the background noise in anechoic conditions. Although the SNRs can be slightly different depending on the reverberant conditions, the input SNRs always stay constant.

To create reverberant conditions, three rooms with different RTs were simulated:

1) A room without reverberation (RT = 0 s),

2) A classroom with desks (RT = 0.5 s; 9.46 m x 6.69 m x 3.00 m), and

3) A restaurant dining area (RT = 1.1 s; 12.19 m x 7.71 m x 2.80 m).

The rooms were modelled in ODEON. The simulated room impulse responses were generated using high-order Ambisonics which was then reduced to third-order Ambisonics to account for the limited number of channels. Target position (0°) and maskers' positions (90, 150, 210 and 270°) were simulated by convolving the appropriate simulated room impulse responses with each stimulus on each trial.

3.2.4 Procedure

There were 9 different conditions based on SNR (-8, -3, +2 dB) and RT (0, 0.5, 1.1 s) of the sound. Each condition was presented in a sperate block, and each block consisted of 20 sentences (180 sentences in total). In addition to that, each participant went through a training round in the beginning, consisting of 20 sentences with random, but all, conditions.

Each trial started with 3 seconds of background noise which was used as the baseline for EEG analysis (baseline phase). It should be noted that there was no reverberation in the baseline for RT = 0 s, whereas for RT = 0.5 s and RT = 1.1 s the baseline was reverberant. After that the HINT sentences were played in the presence of background noise, during which test subjects were required to attend to the target (listening phase) which lasted between 1.2 s to 1.8 s (mean 1.5 s). After the target sentence was finished, the background noise continued for another 2 seconds during which participants needed to maintain the words they just listened to (maintenance phase). When the background noise was stopped, the participants were instructed to repeat all the words within the sentence (recall phase). The performance accuracy was scored based on the correct words they could repeat (i.e., word scoring for speech intelligibility) The procedure for each trial is illustrated in **Fig. 3.1**.

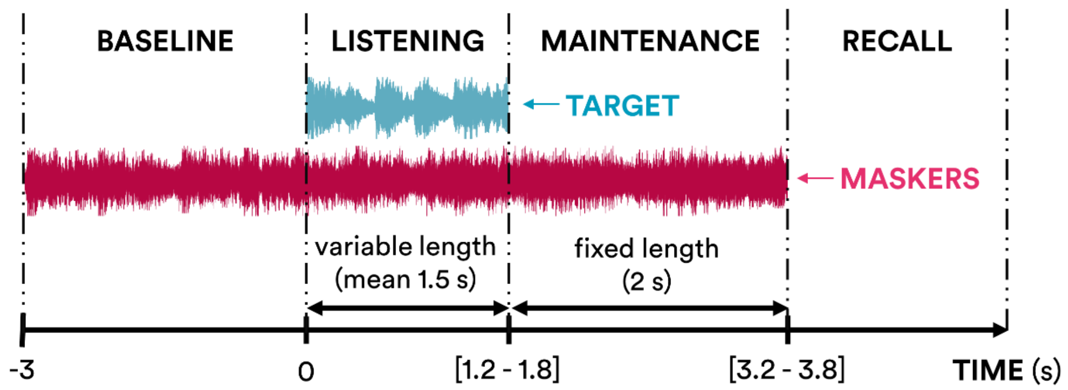


Fig. 3.1. Trial procedure: Each trial started with 3 seconds of background noise (baseline). After that the target sentences were played in the presence of background noise (listening) which lasted between 1.2 s to 1.8 s. After the target sentence was finished, the background noise continued for another 2 s (maintenance). When the background noise was stopped, the participants were instructed to repeat the sentence (recall).

3.2.5 EEG analysis

The processing of EEG data was done similarly to **Section 2.2.5**. On average, 2.1% of the channels were detected as bad and were interpolated. 14.7% of all trials in this experiment were rejected by visual inspection. No participant had more than 24.3% rejected trials.

ERSP window segmentation was followed similarly to **Section 2.2.5.3**. As in **Chapter 2**, The ERSP time window in the baseline was chosen from -2 s to -0.1 s, in the listening phase from 0.1 s to 1.1 s, and in the maintenance phase from 1.3 s to 2.9 s.

Various frequency ranges of power estimation were chosen based on the grand average spectrum. In the listening phase there was a negative peak at 10 Hz, and thus we considered alpha band as 6-13 Hz (see **Fig. 3.3**). In the maintenance phase there were three peaks: a positive peak at 7 Hz, a negative peak at 11 Hz, and a positive peak at 17 Hz. Based on this observation we divided our analyses into three different subgroups: low alpha at 6-9 Hz (see **Fig. 3.4**), high alpha at 10-13 Hz (see **Fig. 3.5**) and beta at 14-20 Hz (see **Fig. 3.6**).

3.2.6 Questionnaire

Participants were asked three questions at the end of each block. The questions were inspired by Zekveld & Kramer (2014), and were translated to Danish. The response to each question had a scale of 0 to 100, with change units of 1.

The first question was "*How many words do you think you understood correctly?*" [English translation] and will be referred as "subjective intelligibility". The second question was "*How much effort did you spend when listening to the sentences?*" [English translation] and will be referred as "subjective effort". The third question was "*How often did you give up trying to perceive the sentences?*" [English translation] and will be referred as "subjective disengagement".

As a post-hoc analysis and based on the obtained results (see **Fig. 3.10**), subjective effort was subtracted from subjective disengagement to obtain a new subjective measure called "derived effort". The idea behind this measure was that the difference between the two measures should show a pattern similar to the EEG. Based on the previous literature (e.g., Zekveld & Kramer, 2014) and the results of current study, participant rate higher subjective effort and higher subjective disengagement when the task becomes more difficult, no matter how close to impossible the task gets. As these measures change in different directions (i.e., one increasing while the other decreasing), I decided to subtract their effect from each other and investigate whether one is greater than the other, depending on the task demand.

3.2.7 Statistics

LMM was used for investigating the effects of SNR and RT on performance, questionnaire, and EEG power in different bands. SNR and RT were fixed factors and participants were random factors in the model. The MATLAB syntax for LMM implementation was $Dependent \sim 1 + SNR * RT + (1|Subject\ ID)$, with *Dependent* being either performance, questionnaire, or EEG power. Both SNR (-5, 0, 5) and RT (-0.53, -0.03, 0.56) levels were centred around zero, for reasons discussed in **Section 2.2.7**.

3.3 Results

3.3.1 Performance

In this experiment, performance accuracy was scored based on the number of correct words repeated within a sentence (i.e., word-based speech intelligibility). In terms of performance accuracy, there were significant effects of SNR ($\beta = 5.98$, SE = 0.30, $t_{158} = 19.67$, $p < 0.001$), and RT ($\beta = -31.17$, SE = 1.78, $t_{158} = -17.49$, $p < 0.001$) and a significant interaction between the two ($\beta = 1.76$, SE = 0.43, $t_{158} = 4.04$, $p < 0.001$). The mean results for performance in each condition is shown in **Fig. 3.2**.

3.3.2 EEG

3.3.2.1 Listening

In the listening phase, there was a negative peak at 10 Hz (alpha ERD). LMM showed significant effects of SNR, whereby a greater SNR related to higher alpha power (less ERD). There was no significant effect of RT and no interaction (**Table 3.1**). The grand average spectrogram, spectrum, topographic map, and alpha power graph chart are shown in **Fig. 3.3**.

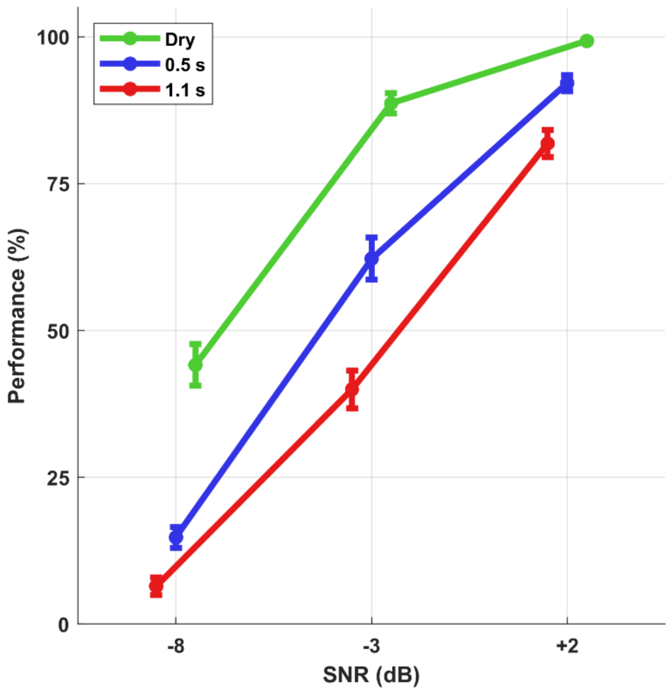


Fig. 3.2 Performance accuracy based on correct repeated words: Increasing SNR and decreasing RT led to higher performance accuracy. Error bars represent standard error of the mean.

Table 3.1 Results of mixed model based on SNR and RT predictors: estimates of relative power changes in the parietal region in different bands and phases. Significant p-values are shown in black font.

DF = 158	Listening	Maintenance		
Band / Predictor	Alpha	Low Alpha	High Alpha	Beta
SNR	**β = 0.50 SE = 0.21 t = 2.36 p = 0.019**	β = 0.07 SE = 0.21 t = 0.35 p = 0.725	**β = 0.87 SE = 0.23 t = 3.73 p < 0.001**	**β = 0.94 SE = 0.20 t = 4.53 p < 0.001**
RT	β = 0.15 SE = 1.95 t = 0.07 p = 0.938	β = 2.36 SE = 1.98 t = 1.19 p = 0.235	β = 3.87 SE = 2.13 t = 1.81 p = 0.072	**β = 6.23 SE = 1.90 t = 3.26 p = 0.001**
SNR x RT	β = -0.42 SE = 0.47 t = -0.89 p = 0.371	**β = -1.33 SE = 0.48 t = -2.73 p = 0.006**	β = -0.54 SE = 0.52 t = -1.04 p = 0.299	β = -0.82 SE = 0.46 t = -1.75 p = 0.081

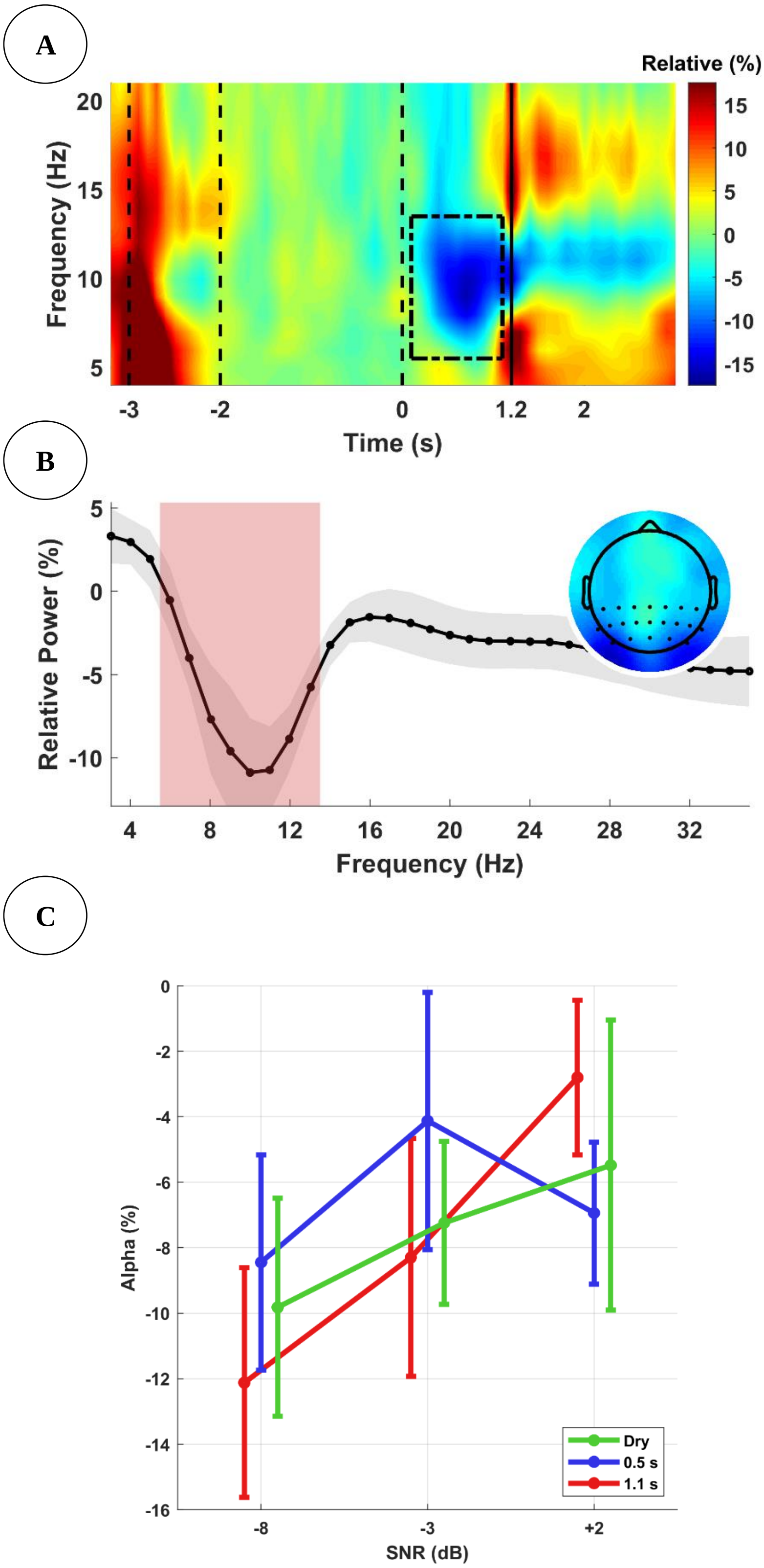


Fig. 3.3 Power changes during listening in the parietal region: A) Grand average spectrogram, B) spectrum and topographic maps in the highlighted time window of panel A of all participants and conditions. C) The modulation of low alpha power by SNR in the highlighted window of panel A which showed significant main effect. Error bars represent standard error of the mean.

3.3.2.2 Maintenance

In the maintenance phase, there was a positive peak at 7 Hz (low alpha ERS), a negative peak at 11 Hz (high alpha ERD), and a positive peak at 17 Hz (beta ERS). Based on **Chapter 2**, the main hypotheses related to the low alpha power for changes of listening effort. LMM results (**Table 3.1**) showed a significant interaction between SNR and RT in low alpha (**Fig. 3.4**). In additional analyses of high alpha and beta, we found a significant effect of SNR in high alpha (**Fig. 3.5**), and a significant effect of both SNR and RT in beta (**Fig. 3.6**).

3.3.3 Questionnaire

The statistical results for subjective intelligibility (**Fig. 3.7**), subjective effort (**Fig. 3.8**), subjective disengagement (**Fig. 3.9**) and derived effort (**Fig. 3.10**) are shown in **Table 3.2**. All the measures show a significant interaction between SNR and RT. The first three measures appear to change linearly by SNR and RT (in the expected directions). However, the derived effort shows higher values for 1.1 s RT in the +2 dB SNR and higher values for dry condition in the -8 dB SNR (i.e., inverted U-shaped pattern).

3.3.4 Performance vs. EEG power

To sort the conditions from the lowest task demand to the highest task demand, the 9 conditions of the experiment were ordered based on the performance (from highest to lowest). Low alpha power in the maintenance phase revealed an inverted U-shaped pattern across these conditions (**Fig. 3.11**). The peak of the inverted U occurred around 60% speech intelligibility.

Using a skipped Pearson correlation test (**Table 3.3**), there was a significant correlation between performance and high alpha as well as beta power in the maintenance phase (**Fig. 3.12**). However, there was no significant correlation between performance and alpha power in the listening phase, or between performance and low alpha power in the maintenance phase.

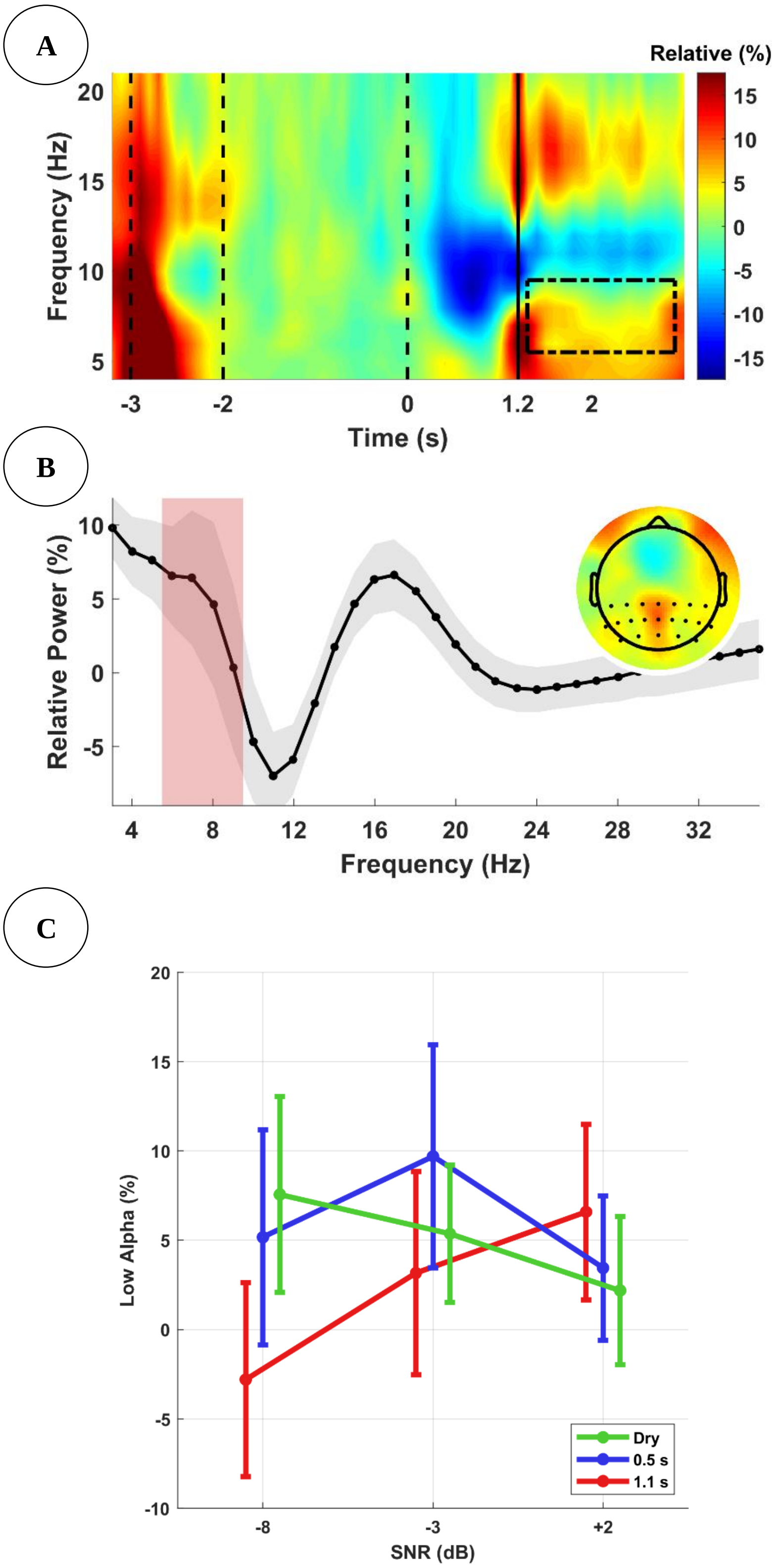


Fig. 3.4 Power changes during listening in the parietal region: A) Grand average spectrogram, B) spectrum and topographic maps in the highlighted time window of panel A of all participants and conditions. C) The modulation of low alpha power by SNR and RT in the highlighted window of panel A which showed significant interaction effect. Error bars represent standard error of the mean.

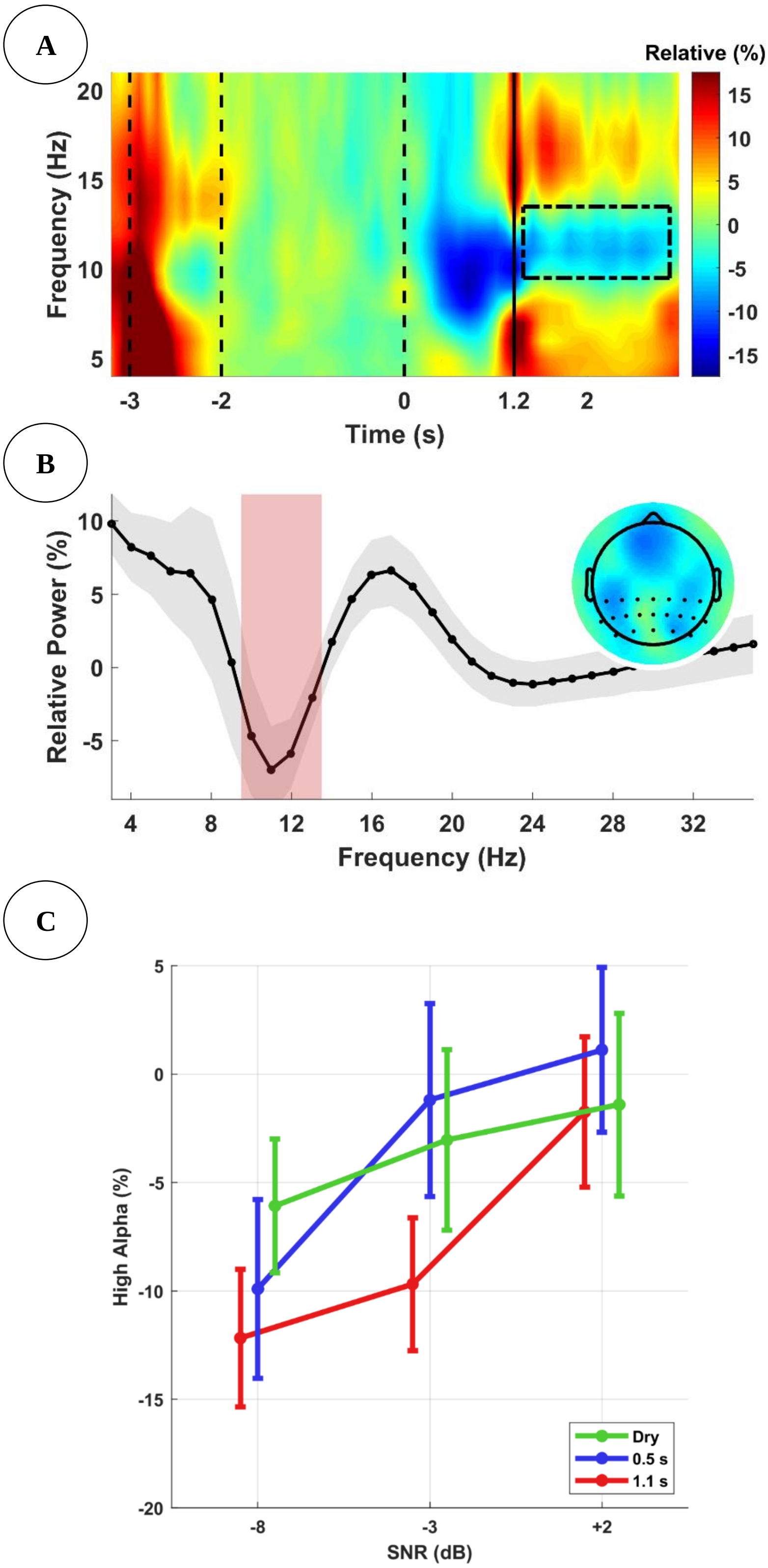


Fig. 3.5 Power changes during maintenance in the parietal region: A) Grand average spectrogram, B) spectrum and topographic maps in the highlighted time window of panel A of all participants and conditions. C) The modulation of low alpha power by SNR and RT in the highlighted window of panel A which showed significant interaction effect. Error bars represent standard error of the mean.

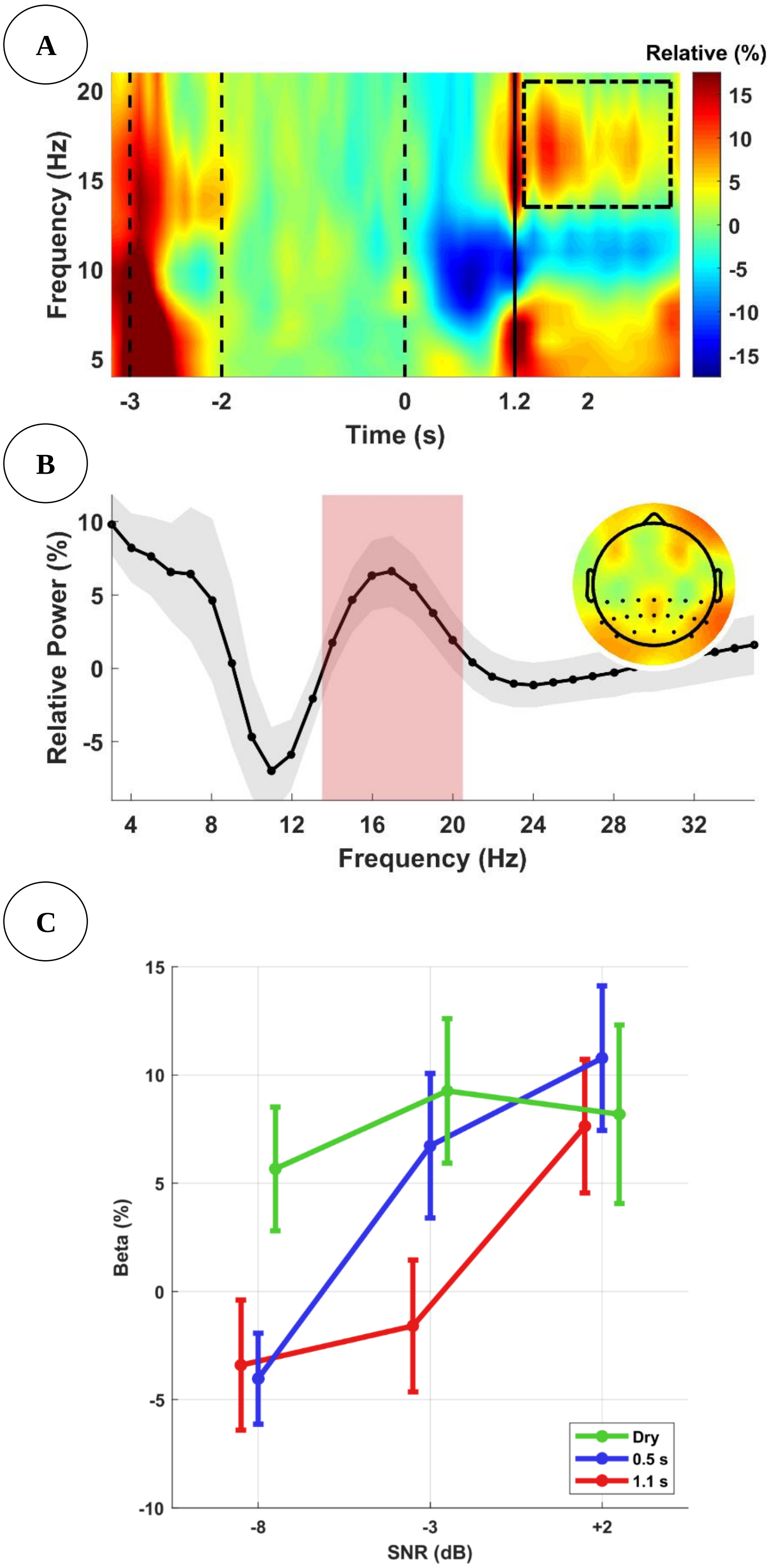


Fig. 3.6 Power changes during maintenance in the parietal region: A) Grand average spectrogram, B) spectrum and topographic maps in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR and RT in the highlighted window of panel A which showed significant main effects. Error bars represent standard error of the mean.

Table 3.2 Results of mixed model based on SNR and RT predictors: estimates of the questionnaire. Significant p-values are shown in black font.

DF = 158	Self-report scales			
Question / Predictor	Subjective intelligibility	Subjective effort	Subjective disengagement	Derived effort
SNR	β = 5.71 SE = 0.42 t = 13.48 p < 0.001	β = -5.60 SE = 0.41 t = -13.57 p < 0.001	β = -5.78 SE = 0.48 t = -11.85 p < 0.001	β = 0.18 SE = 0.58 t = 0.31 p = 0.751
RT	β = -33.74 SE = 2.47 t = -13.61 p < 0.001	β = 23.58 SE = 2.41 t = 9.76 p < 0.001	β = 33.39 SE = 2.85 t = 11.68 p < 0.001	β = -9.81 SE = 3.43 t = -2.85 p = 0.004
SNR x RT	β = 1.56 SE = 0.60 t = 2.57 p = 0.010	β = 1.50 SE = 0.59 t = 2.54 p = 0.012	β = -2.06 SE = 0.69 t = -2.94 p = 0.003	β = 3.56 SE = 0.84 t = 4.23 p < 0.001

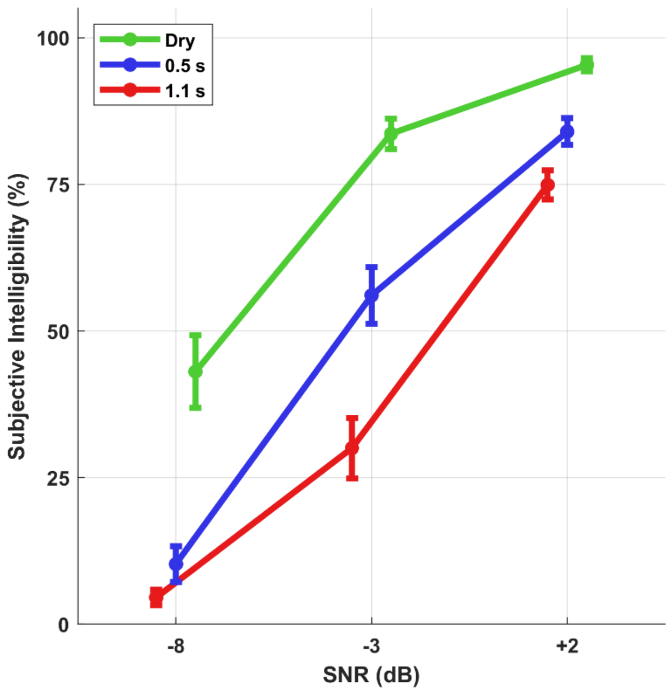


Fig. 3.7 Subjective intelligibility: Increasing SNR and decreasing RT led to higher subjective estimation of intelligibility. Error bars represent standard error of the mean.

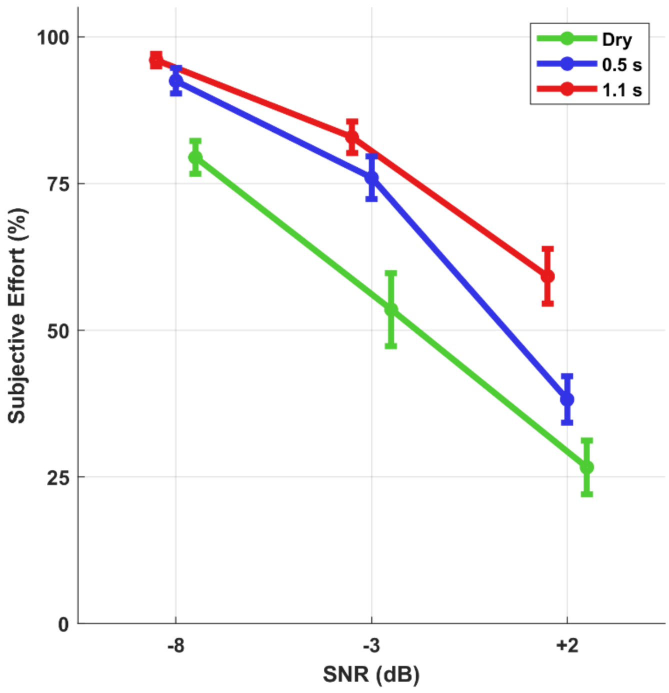


Fig. 3.8 Subjective effort: Decreasing SNR and Increasing RT led to increased perception of effort. Error bars represent standard error of the mean.

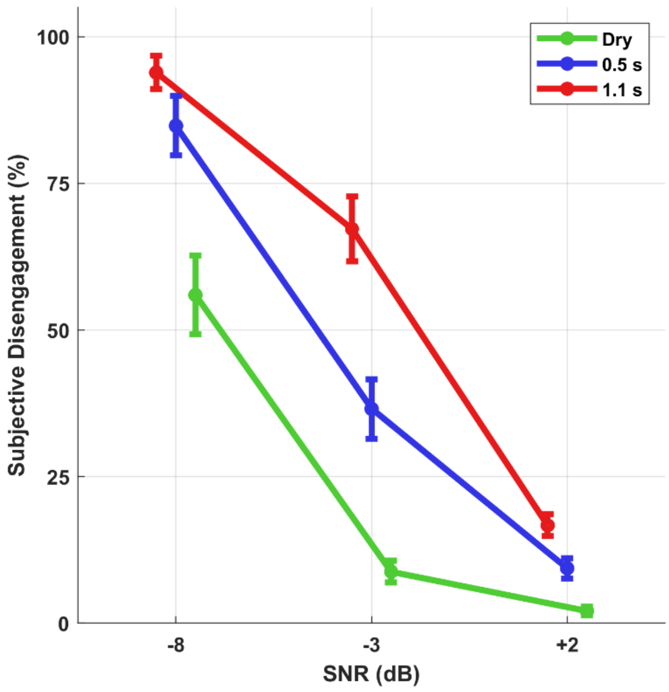


Fig. 3.9 Subjective disengagement: Decreasing SNR and Increasing RT led to increased perception of disengagement from the task. Error bars represent standard error of the mean.

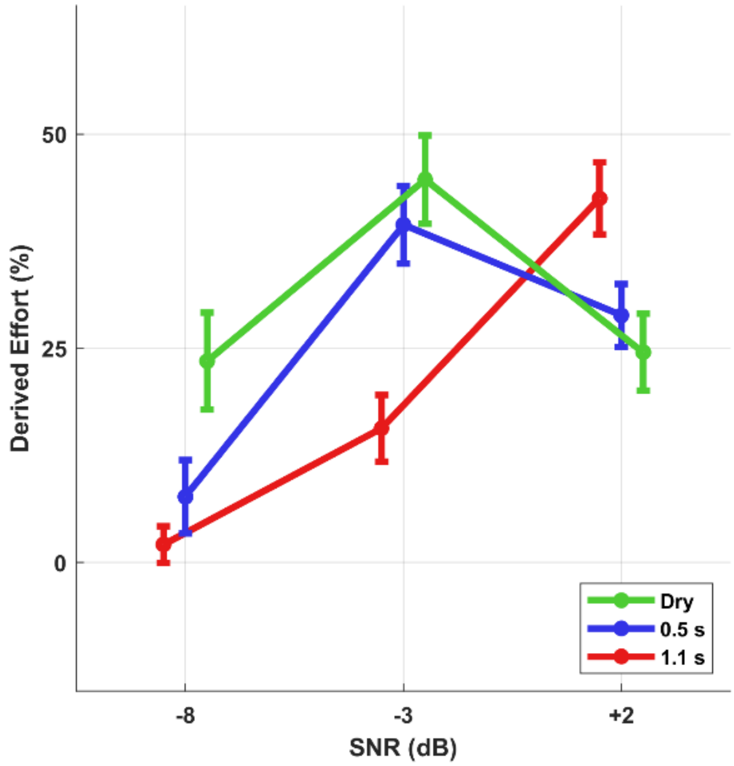


Fig. 3.10 Derived effort: A post-hoc measure derived from subtraction of subjective effort from subjective disengagement which showed interaction between SNR and RT. Error bars represent standard error of the mean.

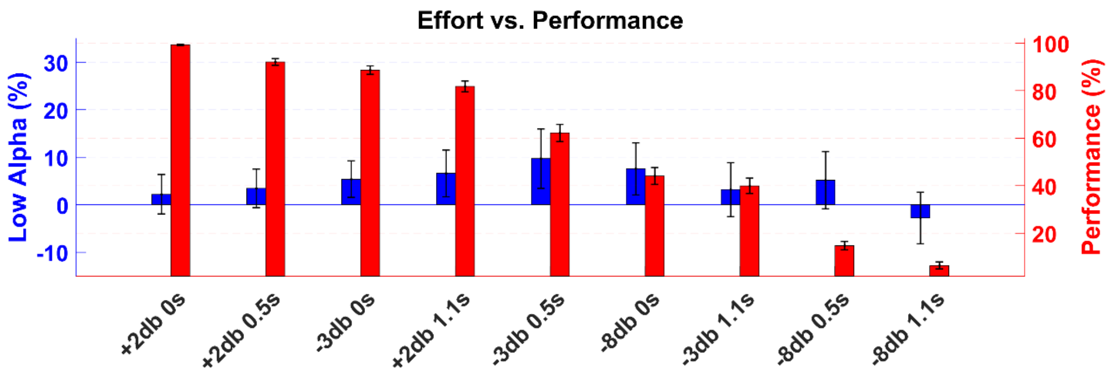


Fig. 3.11 Ordering 9 conditions of the experiment based on the performance (red chart). The changes of low alpha power in the maintenance phase (blue chart) showed an inverted U-shaped pattern.

Table 3.3 Pearson skipped correlation between performance and EEG power in the parietal region in different bands and phases. Significant correlations are shown in black.

	Listening	Maintenance		
Band / **Electrodes**	**Alpha**	**Low Alpha**	**High Alpha**	**Beta**
Parietal	r = 0.12 CI = [-0.04 0.28]	r = 0.14 CI = [-0.02 0.28]	**r = 0.24** **CI = [0.10 0.38]**	**r = 0.33** **CI = [0.20 0.45]**

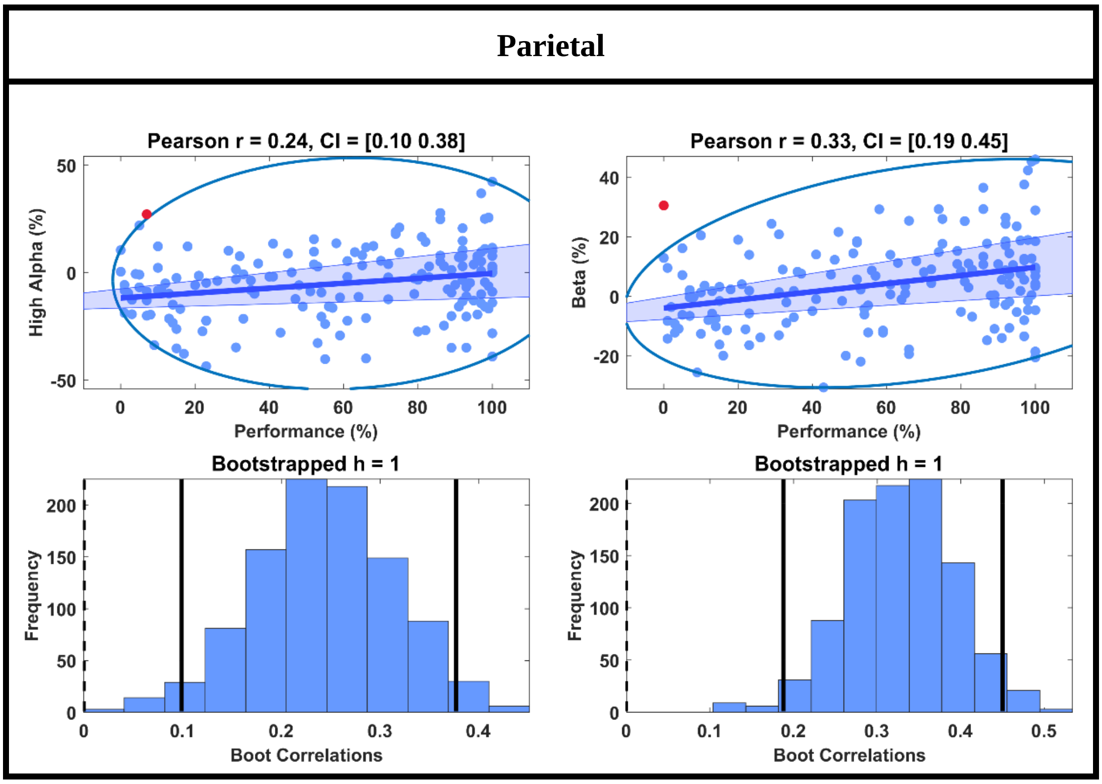


Fig. 3.12 Pearson's skipped correlation with bootstrapping between performance and high alpha and beta power during the maintenance phase in the parietal region. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

3.3.5 Performance vs questionnaire

Based on **Fig. 3.7**, **Fig. 3.8**, and **Fig. 3.9**, self-report scales appear to have changed linearly with SNR and RT across conditions, similar to **Fig. 3.2**. In fact, all three self-report scale questions were highly correlated to performance. Pearson skipped correlation revealed a significant r coefficient between performance and subjective intelligibility (r = 0.95, CI = [0.93 0.96]), subjective effort (r = -0.79, CI = [-0.84 -0.74]), and subjective disengagement (r = -0.94, CI = [-0.96, -0.92]). Derived effort (post-hoc analysis) was also significantly correlated to performance (r = 0.54, CI = [0.44, 0.64]).

Subjective effort and derived effort were also compared to low alpha power in the maintenance phase (as the objective marker of effort). Based on skipped Pearson, subjective effort was not correlated to low alpha power in the maintenance phase (r = 0.01, CI = [-0.14, 0.17]). However, derived effort and low alpha power in the maintenance phase were significantly correlated (r = 0.30, CI = [0.15, 43]; **Fig. 3.13**).

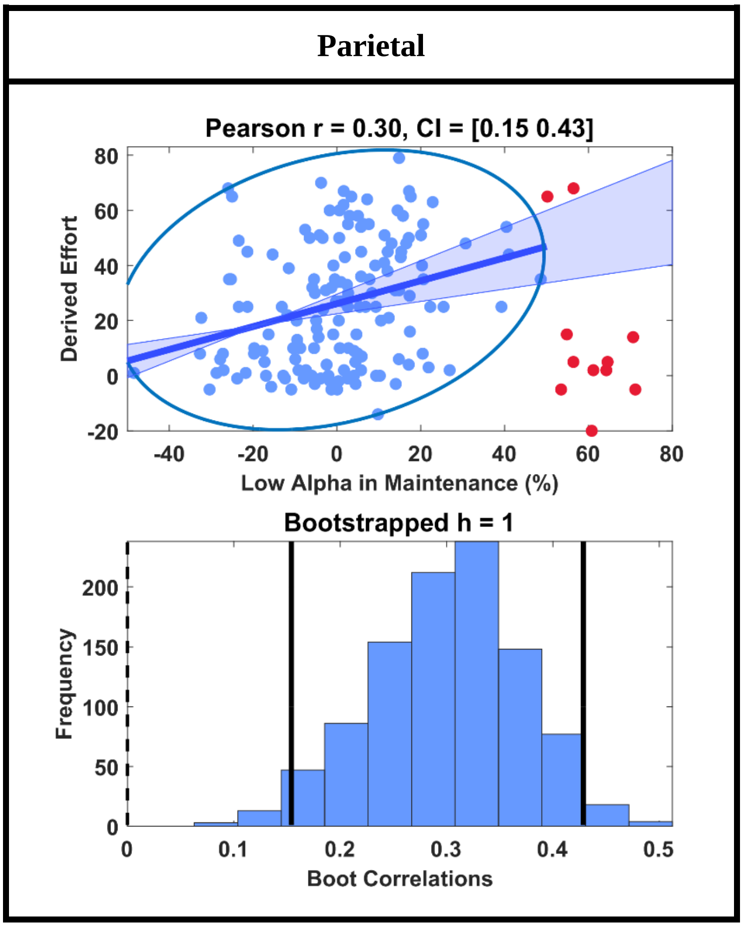


Fig. 3.13 Pearson's skipped correlation with bootstrapping between low alpha power in the maintenance phase and derived effort. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

3.4 Discussion

3.4.1 Overview

In this study, three different SNRs (-8, -3, +2 dB) were used in three different simulated rooms (with RTs of 0, 0.5, 1.1 s) in an anechoic chamber. Decreasing SNR and increasing RT led to more demanding listening conditions that led to lower speech intelligibility in the participants. A questionnaire was used as a subjective measure of effort and EEG low alpha power in the maintenance phase was used as an objective measure of effort. All three questions within the questionnaire were strongly correlated (either positively or negatively) to the speech intelligibility of the participants and significantly changed with both SNR and RT and the interaction between them. Furthermore, low alpha in the maintenance showed an interaction between SNR and RT. Low alpha gradually increased from ~100% speech intelligibility to ~60% speech intelligibility and then started to decrease for the lower speech intelligibilities (i.e., when participants likely started to disengage from the task).

3.4.2 Low alpha power and effort

In **Chapter 2**, it was shown that low alpha power in the maintenance phase revealed a quadratic interaction between SNR and monetary reward. This was in line with the hypothesis of that study which was low alpha power in the maintenance might be a neural correlate of listening effort in a short-sentence paradigm. In the current study, a similar design was used, this time to test whether reverberation (via room simulations) also influences listening effort, and thus, I postulated that low alpha in the maintenance might reflect listening effort modified by SNR and RT.

In fact, low alpha power in the maintenance showed an interaction between SNR and RT. In the dry (RT = 0 s) condition, decreasing SNR led to increased low alpha power in the maintenance phase, whereas in the highly reverberant condition (RT = 1.1 s), decreasing SNR led to decreased low alpha power in the maintenance phase. For RT = 0.5 s, low alpha followed an inverted U-shaped pattern with changing SNRs. To investigate whether there was a pattern to the changes of low alpha with manipulation of SNR and RT, the 9 conditions of the study were ordered from the highest speech intelligibility condition (+2 dB SNR, RT 0 s) to the lowest speech intelligibility condition (-8 dB SNR, RT 1.1 s). Visualizing low alpha power in the maintenance phase with the same order revealed an inverted U-shaped pattern across the conditions (although this could not be statistically proven with LMM, as SNR and RT have different units). The peak of alpha power occurred at ~60% speech intelligibly (-3 dB SNR, 0.5 s RT). Similar findings were also observed in previous studies, where the peak of the inverted U occurred at 60-80% speech intelligibility with another alpha power measure (Decruy et al., 2020), and at 50% speech intelligibility with a pupil dilation measure (Wendt et al., 2018).

Two conclusions can be drawn from the observation that low alpha showed an inverted U-shaped pattern with task demand. The first is that similar to **Chapter 2**, low alpha power in the maintenance phase may be a neural correlate of listening effort. That is, when listening condition was demanding, and more resources in the brain were needed to inhibit task-irrelevant areas of the brain, low alpha power increased. However, when the task became too difficult, and participants started to disengage from the task, low alpha power dropped. The second conclusion is that by looking at speech intelligibility, it might be plausible to predict when the peak of inverted U (whether alpha power or any other objective measure of effort) occurs. According to our data, this happened at ~60% speech intelligibly (**Fig. 3.11**; -3 dB SNR, 0.5 s RT). However, it should be noted that the peak of the inverted U could have happened anywhere between ~80% (+2 dB SNR, 1.1 s RT) to ~45% (-8 dB SNR, 0 s) speech intelligibility. Nonetheless, based on previous literature (Decruy et al., 2020; Wendt et al., 2018), it is plausible that the peak of inverted U occurs around the middle point of psychometric function at ~50% speech intelligibly.

3.4.3 Other EEG indicators

Other than low alpha power in the maintenance phase, significant effects of SNR on alpha during listening (ERD) and high alpha and beta during maintenance (ERS) were observed. In all cases, higher SNR led to increased power (either ERD or ERS). No effects of SNR (-8, -4, 0, +4 dB) in **Chapter 2** was observed. However, there is one key methodological difference between the two studies. While the baseline in the study of **Chapter 2** was kept constant across different conditions, the baseline in the current study varied slightly depending on the reverberation. For example, there was no reverberation in the baseline for RT = 0 s, whereas for RT = 1.1 s the baseline was reverberant. I will discuss in the limitations of the studies that any change in reverberation could slightly change the SNR received by the listener. Given that any EEG power bands in the baseline were averaged together for obtaining ERSP values, it is possible some of the effects of SNRs observed in the results were due to the difference in baselines.

3.4.4 Subjective and objective effort

When asked about subjective intelligibility, subjective effort and subjective disengagement, participants perceived that higher SNR and lower RT led to greater speech intelligibility, less subjective effort and less subjective disengagement. In fact, all three measures were highly correlated (either positively or negatively) to the speech intelligibility of the participants. That means if participants could hear the speech well, they perceived that they spent less effort and were less disengaged from the task.

However, when analysing low alpha power in the maintenance phase (as an objective measure of listening effort), it was the interaction between SNR and RT that shaped an inverted U-shaped pattern of listening effort. The results showed that subjective effort and objective effort were not correlated to each other and showed completely different patterns. While several studies have shown that subjective and objective measures of effort can be correlated to each other, it can be speculated that these measures should not be correlated when the task covers a wide range of demands. In the current study, as changes of the task demand varied speech intelligibility from ~100% to ~5%, subjective and objective effort were not correlated to each other.

Using a post-hoc measure called derived effort, subjective effort was subtracted from subjective disengagement to show if participants' perception of disengagement would offset their perception of effort and whether one is greater than the other in different task demands. While this measure was still correlated to performance (expectedly, as the two measures leading to derived effort were both highly correlated to performance), it was also correlated to low alpha power in the maintenance phase (objective effort). One particular problem with this post-hoc analysis was that even though both subjective effort and disengagement were rated from 0 to 100, but the nature of the questions were different from each other because they required different senses of perception. This can falsify any attempt to put them in the same algebraic equation. While these post-hoc findings should be regarded with great caution, the idea of combining two questionnaires that can closely follow objective measures can be further explored in future studies.

3.4.5 Limitations

One of the methodological limitations of this study was the use of third-order higher Ambisonics which led to a narrow “sweet spot” in the centre of loudspeaker arrays. Sweet spot is a spatial bubble of head location in which the listener would perceive the simulated sound the way it is intended to be heard. The narrow sweet spot called for an accurate placement of participants’ head in the centre of the field. Even slight displacement of head could lead to huge change of speech intelligibility and therefore, listening effort. This problem was mitigated by placing and monitoring a fixed laser pointer on participants’ ears. Other rendering methods can be used in future for obtaining a wider sweet spot.

Another acoustic limitation was how the SNR was defined in this study. As mentioned in **Section 3.2.3**, SNR was defined as the input SPL of target speaker compared to the SPL of background 4-talker babble. That means, the actual sound SNR that reached the participants were slightly different in reverberant conditions. Therefore, this cannot be denied that part of the effects caused by RT were, in fact, slight changes of SNR.

It should also be noted that in real life in high reverberant situations speakers tend to change the way they speak as listening is more difficult compared to low reverberant situations. For example, they might change their speaking pace or pronunciation for better intelligibility. Therefore, in real life listening effort may be ameliorated by adjustments to speech behaviour.

The post-hoc measure introduced as derived effort comes with caveats. Derived effort was extracted from two questions (subjective effort and disengagement) which may not be part of the same domain. That is, when asking about "effort" per se, participants may or may not consider "disengagement" in their answers. Therefore, theoretically, the measure must be interpreted carefully.

3.5 Conclusion

In this study, three different SNRs (-8, -3, +2 dB) and three different simulated rooms (with RTs of 0, 0.5, 1.1 s) were used to manipulate task demand. Speech intelligibility was highly correlated to questionnaires (subjective intelligibility, subjective effort, subjective disengagement), as all of them showed main effects and interaction of SNR and RT. However, low alpha power in the maintenance, as a neural correlate of listening effort, only revealed an interaction between SNR and RT and an inverted U-shaped pattern which was highest at ~60% speech intelligibility. This study showed that reverberation inside a room has an impact on both speech intelligibility and listening effort.

Chapter 4
Listening effort during continuous speech

Two studies in this chapter have been published in peer-reviewed journals. The first study was published by PLOS ONE in 2020:

Seifi Ala, T., Graversen, C., Wendt D., Alickovic, E., Whitmer, W. M., Lunner, T. (2020). An exploratory Study of EEG Alpha Oscillation and Pupil Dilation in Hearing-Aid Users During Effortful listening to Continuous Speech. PLOS ONE 15(7): e0235782. https://doi.org/10.1371/journal.pone.0235782

The second study was published by Ear and Hearing in 2021:

Fiedler, L., Seifi Ala, T., Graversen, C., Alickovic, E., Lunner, T., Wendt, D. (2021). Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise — a combined pupillometry and EEG study. Ear and Hearing - Volume Publish Ahead of Print - doi: 10.1097/AUD.0000000000001050

The following chapter does not include any of the same tables or figures as the publications, with several sections rewritten. Also, despite having behavioural, EEG, and pupillometry measures in both studies, this chapter only, focuses on behavioural and EEG data, as TSA's role was not to analyse the pupillometry data.

It should be mentioned that the EEG analysis in this chapter is slightly different than those presented in the published papers. This decision was made to ensure that the analyses across all chapters of my thesis were consistent and directly comparable to each other.

4.1 Introduction

4.1.1 Continuous speech

In **Chapter 2** and **Chapter 3**, the focus was mainly on the concepts of listening effort and speech intelligibility in short, interrupted speech. While such study designs provide more systematic control over involved parameters during listening, these sorts of stimuli are not typical in everyday life. In real life, most listening situations involve conversations with free-running, continuous discourse, and do not stop after every few words (MacPherson & Akeroyd, 2013; Speaks et al., 1972). For this reason, in this chapter, the focus is on long, uninterrupted speech in the presence of competing talker and background noise to simulate a semi realistic conversation.

A key difference between listening to a short versus long speech is that most of the observed oscillations in the brain during a short sentence are evoked potentials. The duration of event-related activities can vary from few hundreds of milliseconds to 1-2 seconds and can be referred as phasic response. However, using longer stimuli opens a new opportunity for investigating the brain oscillations that are ongoing and more stable and can last longer than phasic response which are referred as tonic response. While phasic response might reflect goal-driven activity, tonic response is more of a sustained activity over a longer period (Dockree et al., 2007). When studying listening effort, it is important to determine if sustained effort is manifested through EEG oscillations (whether alpha or any other bands) during a continuous stimulus (whether speech or non-speech).

Few EEG and pupillometry studies have investigated listening effort during long speech or non-speech stimuli. In one pupillometry study by Zhao et al., (2019), a number of concurrent and spectrally distinct tone streams were presented to young and older populations. Participants were asked to detect gaps in one of the streams in the presence of other disruptive tones. The number of streams (1, 2 or 3) determined the task demand. Each trial lasted for 25 s and pupillometry was recorded as a measure of sustained attention. More distracting streams led to more task difficulty and decreased the performance accuracy of the participants. The results of pupil dilation also were very similar to the results of studies with short stimuli designs (albeit usually using speech stimuli); harder tasks (higher number of streams) led to increased pupil dilation over 25 s of the trial (Zhao et al., 2019).

For EEG studies, however, finding a neural marker for listening effort is not as straightforward due to discrepancy of results between short and long stimuli. In a study by Hjortkjær et al., (2020), participants performed n-back (1-back and 2-back) tasks on the speech sequences with different SNR levels (0 and 10 dB) which lasted for ~50 s. Pupillometry and EEG were recorded together to look for changes in working memory load. Higher load of n-back task and lower level of SNR led to decrease in performance due to increased task demand. While the results of pupillometry was similar to the study by Zhao et al., (2019) (i.e., the harder condition led to increased pupil dilation), posterior alpha power was decreased with higher load of n-back task without any effects of SNR. This observation was in contrast with previous literature using shorter stimuli, that found alpha to increase with higher task demand (see **Chapter 1**). However, they found that speech entrainment (i.e., linear mapping between the speech envelope and the EEG signal) was higher when working memory load was less demanding. This showed that working memory load affects how cortical activity tracks acoustic features of the speech (Hjortkjær et al., 2020), suggesting that the EEG spectrum in such continuous speech paradigm may have been reflecting acoustic features.

The notion that alpha power may not reflect listening effort during continuous speech has also been suggested in two MEG studies by Hauswald et al., (2020). In both studies, participants listened to a continuous speech (approximately lasted between 30 s – 3 mins) and were asked to choose from two nouns that had occurred within the last four words of the trial. In study 1, three different levels of noise vocoding (original, 7 and 3 channels) and in study 2, three additional vocoding levels (5-channel, 2-channel and 1-channel) were also implemented. In both studies lower vocoding levels led to decreased performance. Both studies showed that low-frequency oscillations (1–7 Hz) in frontal regions showed an inverted U-shaped pattern for speech tracking (i.e., coherence between speech envelope and brain activity) with changing degradation levels. However, both studies showed that alpha power decreased with more degradation of the speech similar to the studies by Hjortkjær et al., (2020) or McMahon et al., (2016) and Miles et al., (2017)[1] which all have used long auditory stimuli.

Such EEG and MEG studies on long speech may indicate that the changes of theta and alpha power may reflect acoustic features of the stimuli instead of listening effort. A reasonable consequence of better acoustic during listening is better tracking of the sound. In fact, most of the EEG studies on long speech have shown improvement of speech tracking and attention decoding with better acoustics (e.g., Alickovic et al., 2020; Das et al., 2018; O'Sullivan et al., 2015; Petersen et al., 2017). The difficulties in associating EEG oscillations to listening effort (or speech tracking) during continuous speech points to the importance of these studies. As mentioned earlier

[1] For a detailed description of the studies by (McMahon et al., 2016) and (Miles et al., 2017) see **Section** Error! Reference source not found..

continuous speech have more ecological relevance compared to single sentences in everyday lives.

4.1.2 Hearing loss and hearing aids

Individuals with hearing loss may suffer from a variety of challenges in listening situations including difficulties in speech perception, which can lead to communication difficulties and social isolation (Mathers et al., 2001). In particular, when the listening situation is difficult (e.g., when there is background noise), speech recognition is increasingly challenging for individuals who are hard of hearing (Arlinger, 2003). These issues in speech recognition can cause increased cognitive load, which can in turn lead to negative effects such as difficulties in comprehension (Wingfield et al., 2006), recalling the speech (van Engen et al., 2012; Ward et al., 2016), fatigue (Yang Wang, Naylor, et al., 2018) or disengagement from conversations (Jaworski & Stephens, 1998).

Hearing aids are one of the primary devices that are used for helping people with hearing loss. Hearing aids employ digital noise reduction that can reduce unwanted background noise and help listener to understand speech better. Noise reduction can be based on spatial, temporal, and spectral separation of speech and noise. Spatial separation can be obtained with directional microphones that identify sounds from different directions. Temporal separation relies on modulation rate and depth of the sound. Speech fluctuates at slow rate, but high modulation depth while environmental noise fluctuates at fast rate, but low modulation depth. For spectral separation, an estimate of the noise spectrum is subtracted from the noisy speech. However, the problem with spectral separation is that background noise is constantly fluctuating and therefore a priori is needed for estimation of spectrum in real time (Lakshmi et al., 2021).

Prior works have investigated how noise reduction in hearing aids can reduce listening effort for those with a hearing loss. It was discussed in **Chapter 1** that dual-task paradigm is one of the objective measures to evaluate listening effort. Several studies have used dual-task paradigm to assess the benefits of hearing aids on listening effort. For example, in Picou et al., (2013), a dual task of monosyllable word recognition and the visual reaction time was used for assessing listening effort. There were six conditions changing hearing aids (unaided, aided), visual cues (auditory, auditory-visual), and background noise (present, absent). SNR was set to ~60% speech recognition performance. The performance results showed that that hearing aids and visual cues improved speech recognition performance and background noise impaired performance. The results of reaction time showed no effect of visual cues but increased with background noise and decreased with hearing aids which suggested that hearing aids can decrease listening effort.

Using dual-task paradigms, there are also studies that have shown hearing aids may not benefit listening effort equally in certain populations. In the study by Wu et al., (2014), elderly participants (56 to 85 years old) performed a speech recognition task with either a simulation driving task or a visual reaction-time task. Three hearing aid conditions of unaided, aided with omnidirectional processing (OMNI), and aided with directional processing (DIR) were used. The change in the driving task or the visual reaction time were used as a measure of listening effort. In both tasks (driving performance and visual reaction time) speech recognition was higher in the OMNI and DIR conditions than in the unaided condition, but the other dual task was not affected by amplification and directional processing. These results showed that in elderly participants hearing aids may not benefit listening effort. In a follow-up study, the visual reaction-time dual-task experiment was conducted on younger adults with normal hearing. This time the results indicated that the OMNI and DIR conditions significantly improved speech recognition and reaction time. These findings suggests that the benefit of hearing aids on listening effort which are measured from younger, normal-hearing population may not be translated to older, hearing-impaired population.

The benefits gained from hearing aids noise reduction can be dependent on task demand as well as individual differences. For example, the study by Ng et al., (2015) suggest that noise reduction may work more effectively in high task demand situations on people with better working memory capacity. In this stuyd, they used sentence-final word identification and recall (SWIR) and reading span tests on hearing impaired participants. Native speech was used as the target, while competing background babble was either in native or foreign languages. While word-recall performance was ~60% for all conditions, noise reduction schemes in hearing aids only improved recall performance in native (more disruptive) but not foreign competing speech. This might be because hearing aid noise reduction was more effective when listening demand was higher. Also, in the same study, they found that participants with better reading span results (i.e., higher working memory capacity) showed recall benefit of noise reduction across list positions, while participants with worse reading span results (i.e., lower working memory capacity) mainly improved in late list positions. This showed that the effectiveness of noise reduction was dependent on listener's working memory capacity. Even though participants with worse reading span might had heard the words in earlier lists, they could not remember it probably due to high workload on their working memory because of effortful listening.

There are also studies that have used physiological measurements to demonstrate that the benefit of hearing aid processing is impacted by listening demand using. For example, in a pupillometry study by Ohlenforst et al., (2018), a wide range of SNRs (from -16 to 12 dB) with two different masker types (4-talker babble and stationary noise) were used in a sentence recognition task to test the effects of noise reduction scheme. Pupillometry data was simultaneously recorded for assessing listening effort. They observed an inverted U-shape curve in mean and peak pupil dilation across the ranges of SNR. For the 4-talker babble condition, the peak of the curves shifted approximately 5 dB towards lower SNRs when noise reduction scheme was activated. They concluded that with the 4-talker masker, the noise reduction scheme helped participants improve their intelligibility and also delay "giving up" under demanding listening situations. Similar to study by Ng et al., (2015), one type of masker (4-talker babble) disrupted the speech more compared to the other masker (stationary noise) and noise reduction was more effective during the demanding condition.

These studies by Ng et al., (2015) and Ohlenforst et al., (2018) are particularly important since they use maskers that are similar to the conditions that listeners will experience in everyday life (albeit using short-sentence paradigm), showing the effectiveness of hearing aid noise reduction methods in ecologically valid situations. Having different talkers in the background (Ohlenforst et al., 2018), speaking with the same language (Ng et al., 2015), is a common theme of our daily lives' listening situations. However, interestingly, these studies suggest that hearing aids noise reduction is effective *only* during demanding conditions. If the listening conditions are not demanding, then they appear to minimally impact the expenditure of users' listening effort.

Given the early evidence that listening effort is employed quite differently between short and long speech stimuli, it is clear that to fully understand the experience of hearing-impaired individuals in everyday life we cannot simply assume that the findings from prior work using short stimuli extrapolate. There is currently a scarcity of research into continuous speech processing. In this chapter, we begin to address this gap by exploring listening effort in a comparable way to prior chapters, but now using continuous speech stimuli in hearing-impaired individuals.

4.1.3 Objectives

To explore listening effort in hearing-impaired individuals during continuous speech, two different studies are presented in this chapter. The first study was a pilot study in which hearing-impaired participants were exposed to uninterrupted speech in presence of a competing talker and background noise. We investigated how the changes of SNR in a continuous speech affected listening effort, measured by EEG alpha power. The hypothesis was that increasing SNR should decrease task demand which then leads to increased alpha power based on “function inhibition” theory (**Section 1.4.4**).

After gaining insight on how alpha power manifests in continuous speech, the second study tests the benefit of specific hearing device processing by manipulating noise reduction. Similar to the first study, hearing-impaired participants listened to uninterrupted speech in presence of a competing talker and background noise. However, this time SNR and noise reduction scheme (Off vs. On) were manipulated in order to look for any benefits of hearing aids noise reduction on listening effort, measured by EEG alpha power. Similar to the hypothesis for the first study, increased alpha power was expected due to increased task demand in the second study. This time, changing SNR and activating/deactivating noise reduction were task demand manipulations.

4.2 First Experiment

4.2.1 Study design

4.2.1.1 Participants

Eight native Danish-speaking adults with an average age of 70 ± 12 years participated in the study and signed a written consent form prior to study onset. Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark. All test participants were experienced hearing-aid users (at least for 4 months) with symmetrical, mild, sensorineural hearing loss. The pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz was 31 ± 5.5 dB HL. The difference between the left and right ear in air conduction hearing thresholds for 0.5, 1, 2 and 4 kHz were calculated and averaged together. No participant had more than 5 dB average difference between the ears. Participants had no history of neurological disorders, dyslexia or diabetes.

4.2.1.2 Hearing aids

The participants were fitted binaurally with behind-the-ear Oticon Opn1 mRITE hearing aids with miniFit Speaker Unit 85. Domes used in the test corresponded to what the test subject was currently using: either miniFit open domes or miniFit Bass domes with 1.4 mm vent effect. Noise reduction and directional microphones were deactivated so that the hearing aids just provided individualized audibility via the proprietary gain and frequency prescription rule. Volume control and the mute function were also deactivated to prevent the test subjects from changing the gain during testing.

4.2.1.3 Apparatus

The experimental setup consisted of three loudspeakers positioned at ±30° and +180° azimuth relative to the participants (see **Fig. 4.1**). The loudspeakers in the front hemifield were the target and contralateral distractor locations, symmetrically off-center to counterbalance any asymmetrical hearing abilities, and the loudspeaker in the rear hemifield presented 4-talker babble noise to increase task complexity. A computer screen for displaying the instructions and the questions were positioned in front of the participants in a way not to cause acoustic shadowing. The spatial setup of the test is illustrated in **Fig. 4.1**. The descriptions for loudspeakers and EEG device are the same as **Section 2.2.2**.

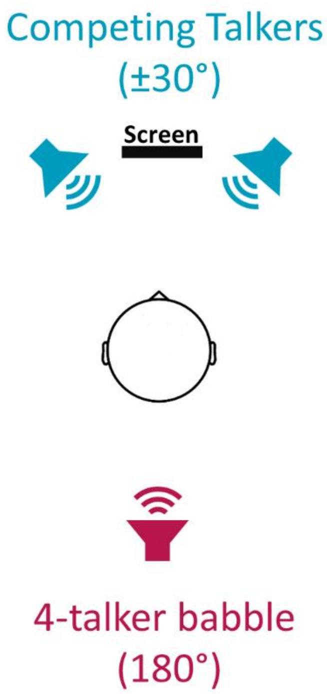


Fig. 4.1 Spatial setup of the task: Three loudspeakers positioned at ±30° and +180° azimuth relative to the participant’s head. The loudspeakers in the front hemifield (in blue) were the target and contralateral distractor locations, and the loudspeaker in the rear hemifield (in red) presented 4-talker babble noise. The loudspeakers were 1.2 m away from the listener.

4.2.1.4 Stimuli

Non-dramatic Danish news clips with neutral content were used for the target and contralateral distractor speech (30 seconds), while the 4-talker babble noise (35 seconds) was provided by Danish audiobooks. The target and distractor speech were read by a randomized male or a female speaker, and for each trial the target and distractor were never the same gender.

The A-weighted SPL at the center of the room was 50 dB for the babble and 65 dB for the target on every trial. The contralateral distractor level was either 65 dB or 70 dB on each trial to generate two different SNR conditions: 0 and -5 dB. For this study, SNR was defined as the long-term average sound level of the target signal (with pauses longer than 200 ms being cut out) compared to the competing front talker only. Although both SNRs were relatively low compared to common environments for hearing-aid wearers (cf. Smeds, Wolters and Rung, 2015), the 0 dB and -5 dB SNR conditions will be referred as "high SNR" and "low SNR", respectively.

4.2.1.5 Procedure

Before each trial, the participants were instructed on the screen to pay attention to the target on the right or left side and ignore the talker on the other side and the babble behind them. The location of the target (i.e., right or left front loudspeaker) was randomized between each trial.

There were 54 trials for each SNR, randomly distributed across all 108 trials. Each trial consisted of 35 seconds of 4-talker babble played in the background. The target and distractor speech were presented 5 seconds after the onset of the babble (i.e., after the baseline period) and then continued for 30 seconds, followed by a three-choice question regarding the content of the attended target audio clip [e.g., "Who warns against the dangers of discrimination?" (English translation)] (See **Fig. 4.2**). Participants were given a rest period every 36 trials (i.e., twice during the experiment), while minor breaks were given between every 8^{th} trial. To acclimatize to the hearing aids, the participants listened to four training trials before starting the experiment.EEG analysis.

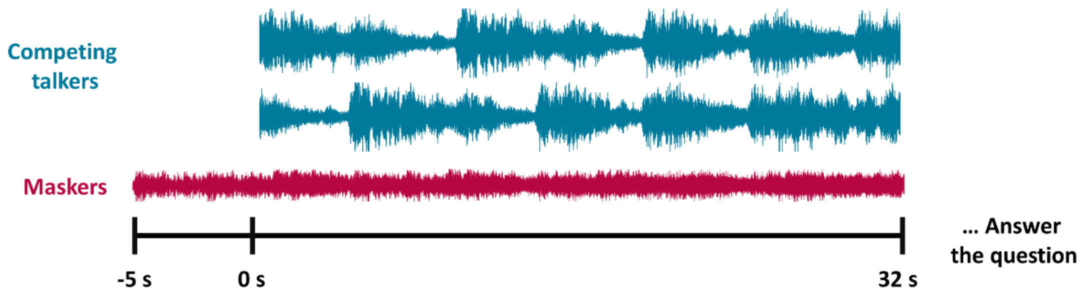


Fig. 4.2 Trial procedure: Each trial consisted of 35 seconds of 4-talker babble played in the background (in red). The target and distractor speech were presented 5 seconds after the onset of the babble and then continued for 30 seconds (in blue), followed by a three-choice question regarding the content of the attended target audio clip.

4.2.1.6 EEG analysis

The pre-processing of EEG data was done similarly to **Section 2.2.5.1**. In total, 1.8% of the channels were detected as bad and were interpolated. Also, 4.7% of all trials in this experiment were rejected by visual inspection. No participant had more than 11.1% of trials rejected.

Power extraction was also done similarly to **Section 2.2.5.3**, except that due to longer length of stimuli, Morlet wavelets (7 cycles width) were centred at 500 ms steps within a trial (frequency range between 2 and 35 Hz). Baseline for analysis of ERSP was -4 to 0 s prior to start of the target. The first second of background noise (-5 to -4 s) was removed to avoid ERPs. The EEG power for target stimuli was divided into two different sub-groups based on the timeline: phasic power (0 to 1 s) and tonic power (1 to 30 s).

For the phasic power, alpha power ERD (similar to listening phase in **Chapter 2** and **Chapter 3**) in the parietal region was investigated. For the tonic power, theta (exploratory), alpha (hypothesis-driven), and beta (exploratory) power in the parietal and fronto-central regions were explored. In this study, theta is defined as 6 – 7 Hz (7 Hz peak in the fronto-central region; see **Fig. 4.5**), alpha as 8 – 12 Hz (10 Hz peak in the parietal region; see **Fig. 4.4**), and beta as 13 – 20 Hz (despite no apparent peak; see **Fig. 4.6**, just to keep comparison between studies more consistent).

4.2.1.7 Statistics

To estimate the effects of SNR on performance and EEG power in different bands, several LMMs were implemented. For this study, SNR was the fixed factor with participants as random factors. The MATLAB syntax for LMM implementation was $Dependent \sim 1 + SNR + (1|Subject\ ID)$, with *Dependent* being either performance or EEG power. The estimates of the LMM (β), t-values and degrees of freedom (t_{DF}) and P values are reported in the Results section.

Due to the longer length of the stimuli, changes of EEG power over time were also investigated. For this purpose, EEG power was averaged in 1s windows, to get a single value per second. Then, a first-degree polynomial curve was fitted to the alpha power in two different ways. The first approach was to fit the curve over the whole trial (30 s) to investigate the slope of EEG power changes. For statistical evaluation, a one-sample t-test was used to look for any positive or negative slope (i.e., testing against 0). The second approach was to break down alpha power to smaller chunks to investigate the time dynamic of slope changes (i.e., were the slope changes most dominant in the beginning or at the end of the stimuli?). The same process (polynomial fit and one-sample t-test evaluation) was repeated for each 5-s period instead of the whole trial (e.g., 0-5 s, 5-10 s, 10-15 s, etc.). To correct for multiple statistical tests, false discovery rate (FDR), using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), was applied.

To explore the correlation between EEG power (theta, alpha, beta) and performance, a Pearson skipped correlation was used, as described in **Section 2.2.7**.

4.2.2 Results

4.2.2.1 Performance

There was a significant improvement in performance due to increased SNR ($\beta = 3.24$, $SE = 0.54$, $t_{14} = 5.92$, $p < 0.001$) reflecting that the participants benefited from higher SNR in terms of understanding the contents of the speech (**Fig. 4.3**). Also, the above chance performance for the low SNR suggests that the speech in the worst condition was still partly intelligible.

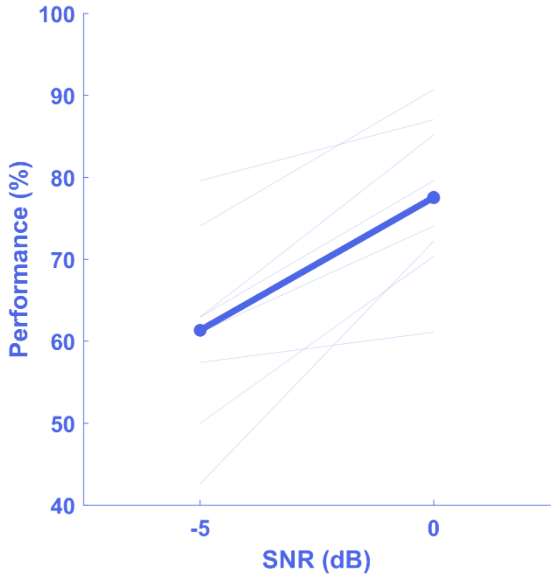


Fig. 4.3 Performance results: The correct percentage of the questions that were answered regarding the contents of the target speech.

4.2.2.2 EEG

4.2.2.2.1 Phasic power

Analysis of EEG power showed desynchronization of phasic alpha power (first second of the stimuli), but it was not significantly modulated by SNR in either the parietal or fronto-cental regions (see **Table 4.1**).

4.2.2.2.2 Tonic power

In the parietal region, tonic theta, showed no significant changes due to SNR. However, tonic alpha (after the first second) showed synchronization throughout the trial that was significantly increased by higher SNR (**Fig. 4.4**). Despite no apparent peak in the beta range, tonic beta also showed the same pattern and was significantly increased with higher SNR (**Fig. 4.6**). The details of the results are in **Table 4.1**.

In the fronto-central region, despite synchronization of tonic theta, there was no significant modulation by SNR in theta (**Fig. 4.5**) or alpha band. However, tonic beta was significantly increased by higher SNR. **Table 4.2** shows the results of the fronto-central region.

4.2.2.2.3 Changes over time

The analysis of changes over time showed no significant effects (either in grand average data (**Fig. 4.7 Top row**) or individual SNR conditions (**Fig. 4.7 Middle row**). The analysis of 5-s slopes across the trial also did not show any significant results after FDR correction (**Fig. 4.7 Bottom row**).

4.2.2.3 Correlations

The analysis of Pearson skipped correlation showed that there was no significant correlation between performance and EEG power in any bands (**Table 4.3**). Even though both performance and tonic alpha/beta were significantly modulated by SNR, there was no correlation between performance and alpha/beta power (**Fig. 4.8**).

Table 4.1 Results of mixed model based on SNR predictor: estimates of different relative power changes in the parietal region in different bands and phases. Significant p-values are shown in black font.

DF = 14	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = -0.56 SE = 0.43 t = -1.31 p = 0.209	β < 0.01 SE = 0.23 t = 0.03 p = 0.972	**β = 0.64 SE = 0.22 t = 2.90 p = 0.011**	**β = 0.57 SE = 0.15 t = 3.69 p = 0.002**

Table 4.2. Results of mixed model based on SNR predictor: estimates of different relative power changes in the fronto-central region in different bands and phases. Significant p-values are shown in black font.

DF = 14	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = -0.68 SE = 0.64 t = -1.06 p = 0.306	β = 0.54 SE = 0.37 t = 1.43 p = 0.172	β = 0.62 SE = 0.33 t = 1.90 p = 0.077	**β = 0.50 SE = 0.17 t = 2.80 p = 0.014**

4.2.3 Summary

The first study showed that alpha and beta power in hearing-impaired participants were increased due to higher SNR in a continuous-speech paradigm. This pilot study showed that the pattern of alpha power in continuous speech under demanding conditions is different than what studies with short-speech paradigms have shown. To gain a better perspective on listening effort in continuous speech, a second experiment was conducted in which a similar paradigm was used with a larger population of hearing-impaired participants. This time the interaction between SNR and hearing aids noise reduction scheme to affect listening effort was investigated.

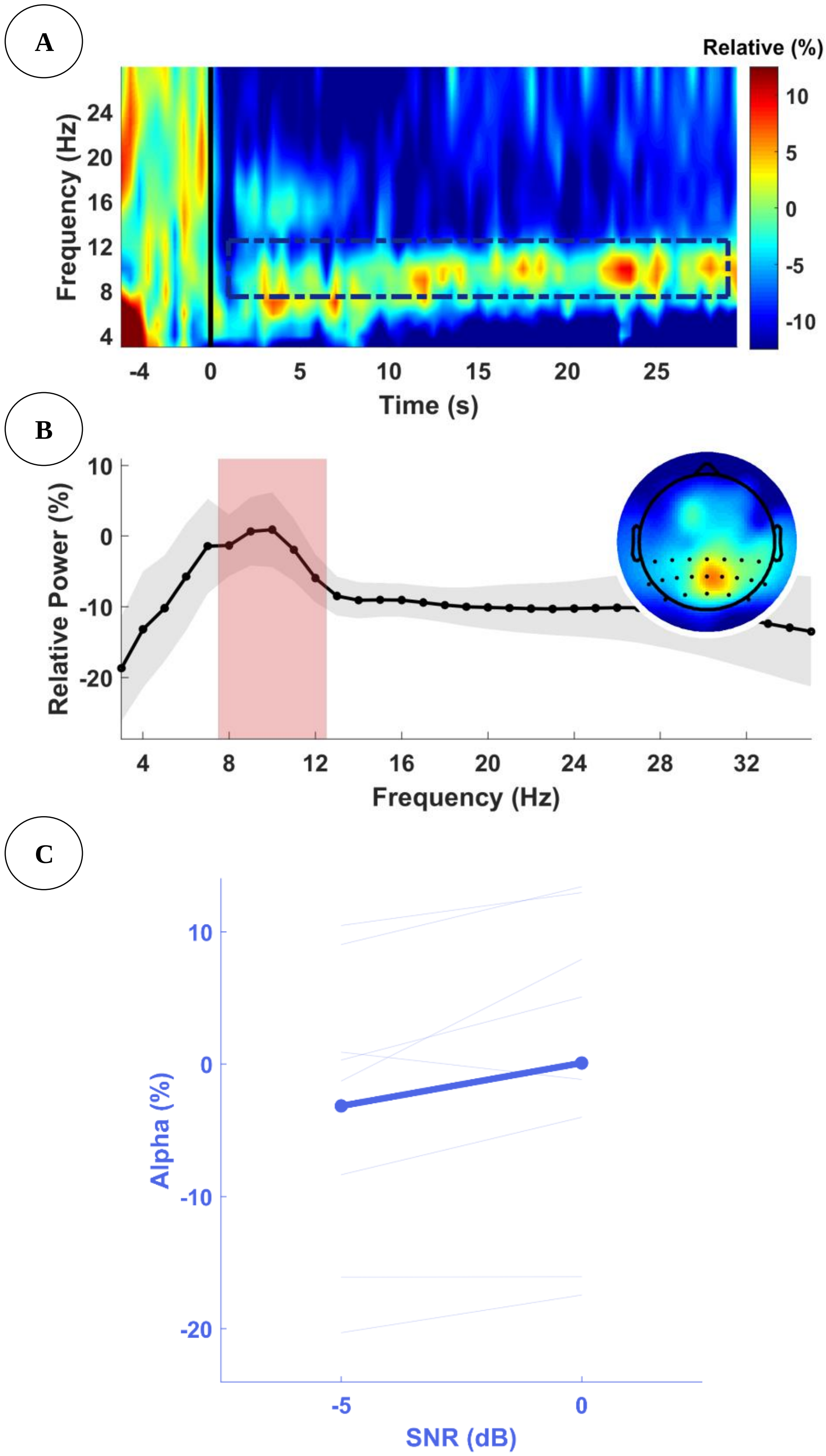


Fig. 4.4 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of alpha power by SNR in the highlighted window of panel A which showed significant effect.

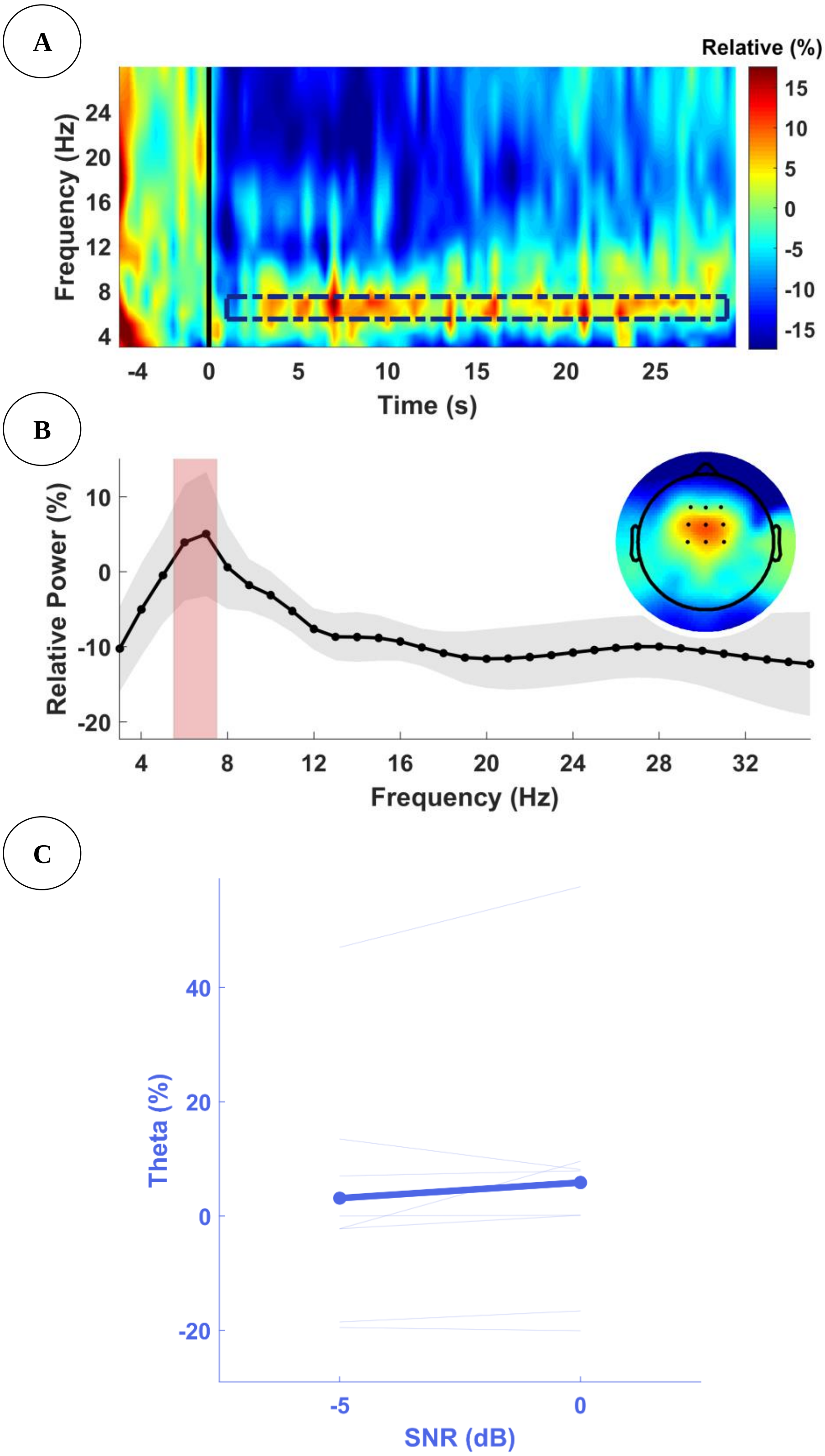


Fig. 4.5 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of theta power by SNR in the highlighted window of panel A which showed no significant effect.

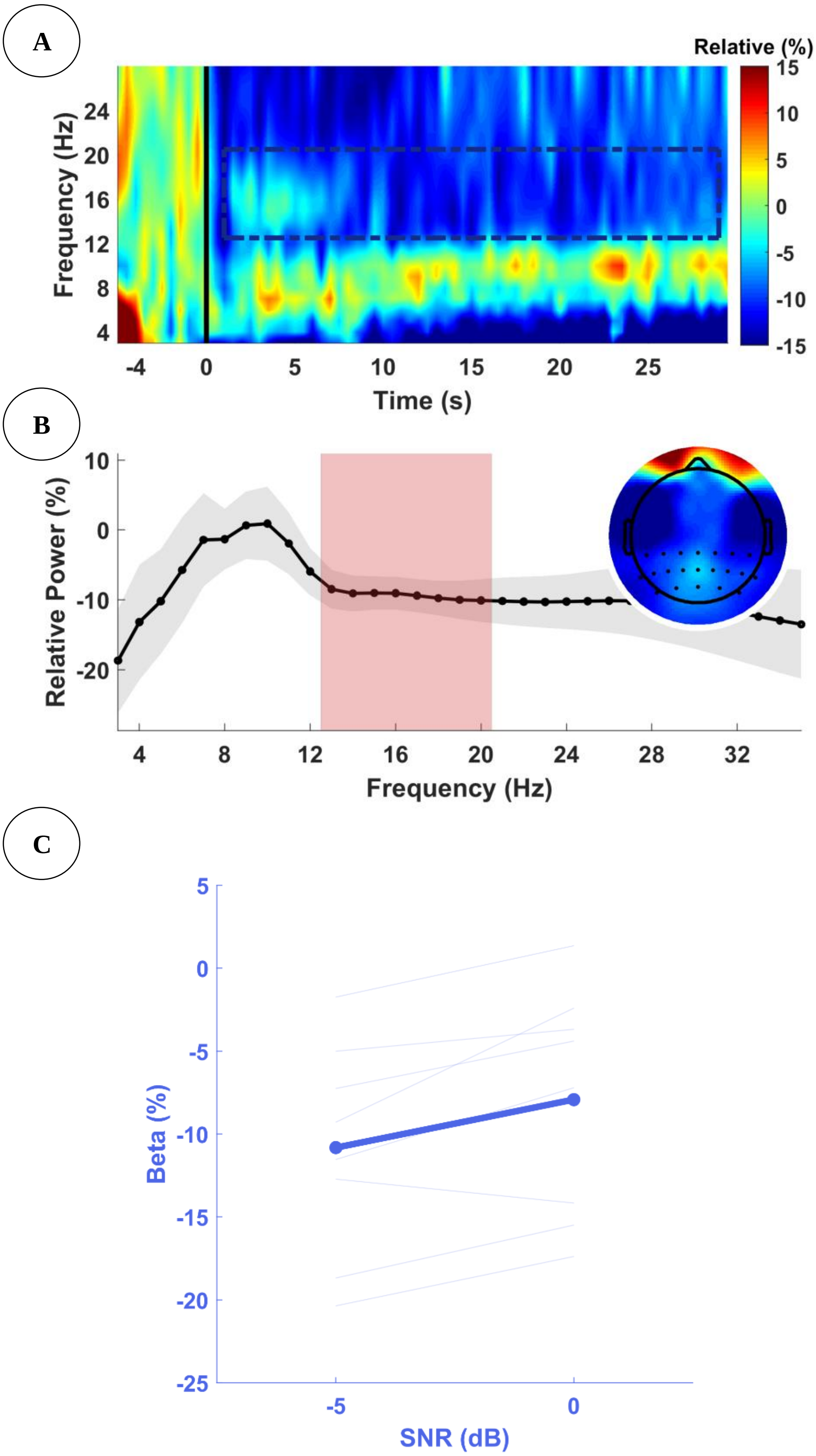


Fig. 4.6 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR in the highlighted window of panel A which showed significant effect.

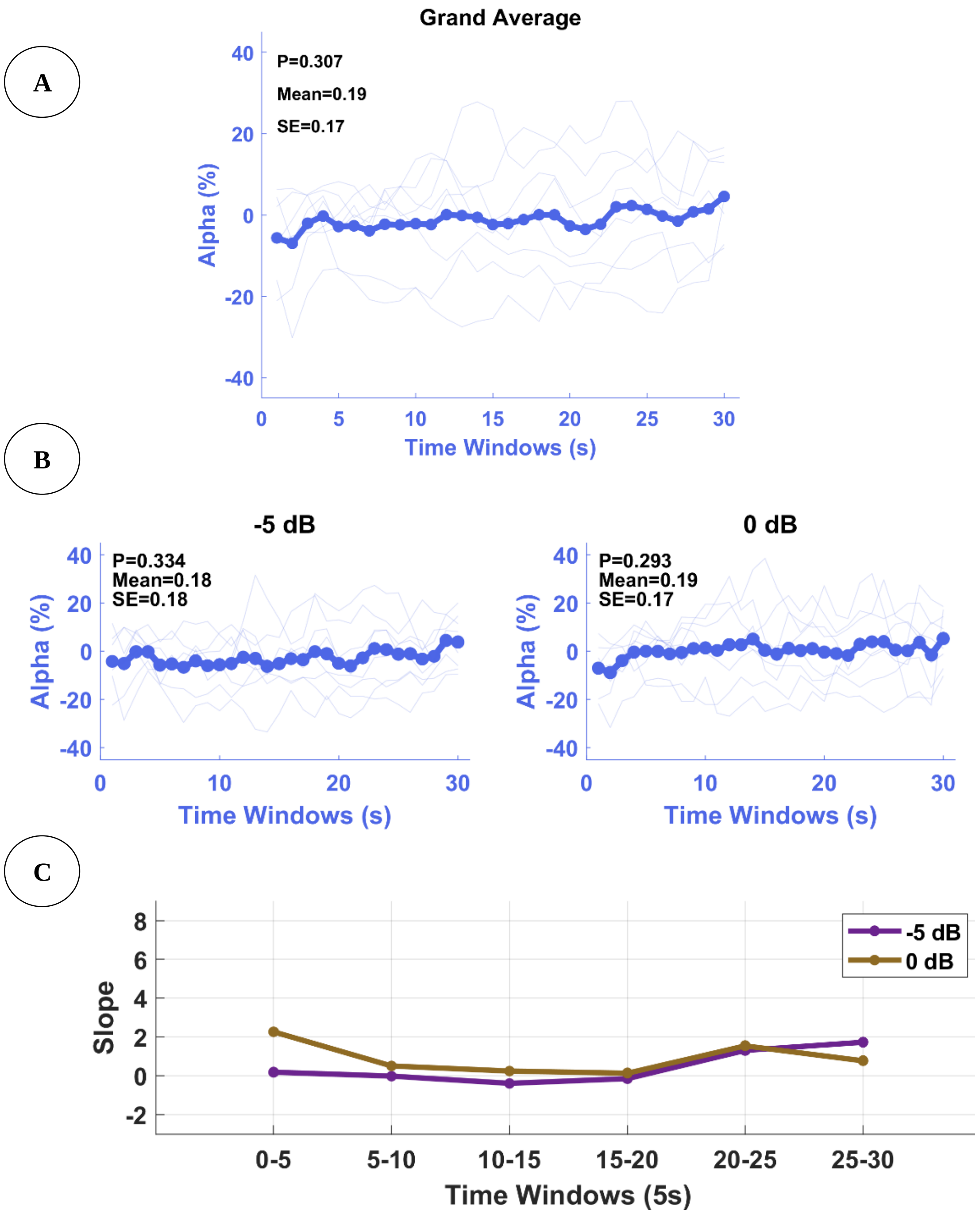


Fig. 4.7 Changes of alpha power over time for A) grand average data and B) each condition in 1-sec windows with no significant slope. C) The analysis of slope within 5-s intervals did not show any significantly positive nor negative values in any time windows.

Table 4.3 Pearson skipped correlation between performance and EEG power in different regions, bands, and phases.

	Phasic	Tonic		
Band / Electrodes	Alpha	Theta	Alpha	Beta
Parietal	r = -0.03 CI = [-0.55 0.61]	r = 0.04 CI = [-0.44 0.59]	r = -0.07 CI = [-0.50 0.49]	r = -0.08 CI = [-0.51 0.47]
Fronto-central	r = -0.11 CI = [-0.63 0.55]	r = -0.07 CI = [-0.49 0.52]	r = 0.07 CI = [-0.39 0.63]	r = -0.20 CI = [-0.61 0.49]

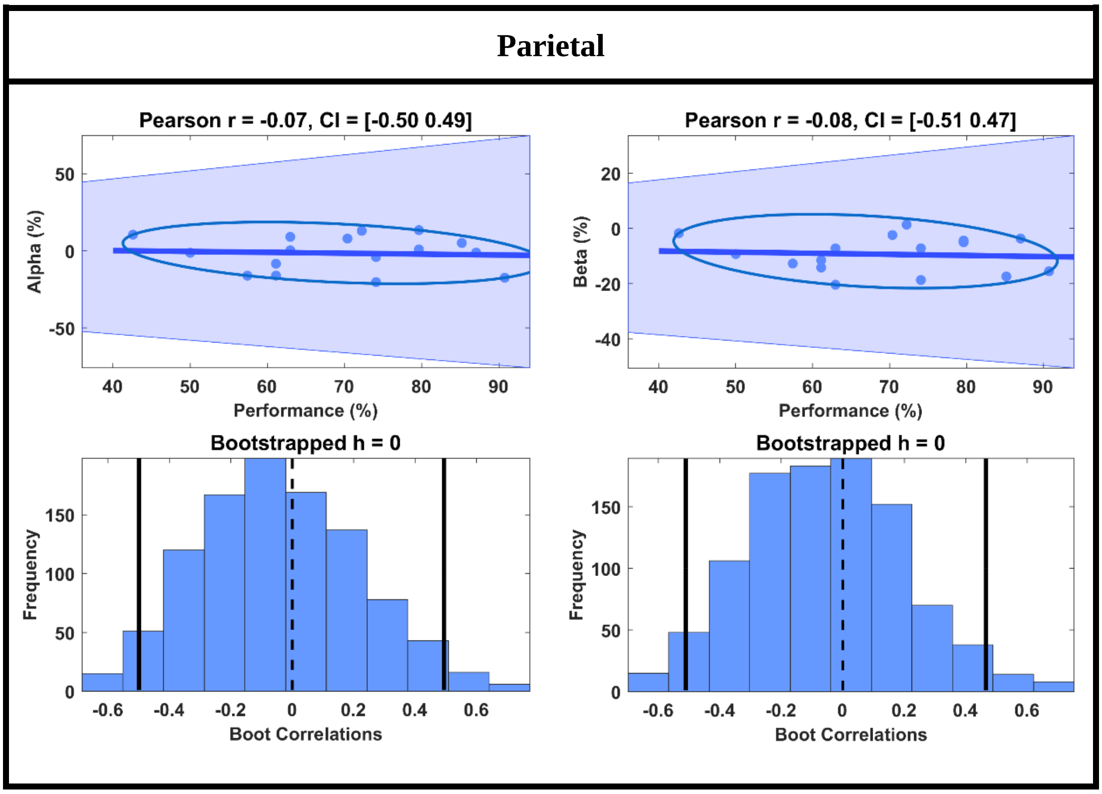


Fig. 4.8 Pearson's skipped correlation with bootstrapping between performance and tonic alpha and beta power in the parietal region. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI. No significant correlation (h = 0) was observed.

4.3 Second experiment

4.3.1 Study design

4.3.1.1 Participants

For this experiment 22 hearing-impaired participants with an average age of 67 ± 11.2 years were recruited. The study was approved by the ethics committee for the capital region of Denmark and all participants signed a written consent prior to the experiment. Participants had at least 4 months of experience with their hearing aids. The pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz ranged from 33 to 58 dB HL. The average difference between the left and right ear in air conduction hearing thresholds for the mentioned frequencies was a maximum of 8 dB. The inclusion criteria were mild to moderate sensorineural hearing loss, and no history of neurological disorders, dyslexia or diabetes.

4.3.1.2 Hearing aids

All participants were fit with identical hearing aids with non-individualized, closed soft tips. Two pairs of hearing aids were fit for each participant. In one pair, noise reduction was turned off and in the other pair, noise reduction was activated. Other than that, all signal processing features (e.g., feedback cancelling) were kept at default and did not vary between the conditions.

Both pairs of hearing aids amplified the sound based on each individual's hearing threshold via the Voice Aligned Compression (VAC) rationale. The VAC amplification rationale is based on a wide dynamic range compression scheme with compression knee points between 20 and 50 dB SPL depending on the frequency range and the individual's hearing thresholds. Similar to other standard fitting procedures (Keidser et al., 2011), VAC was developed to fit hearing aids to individual needs to improve overall speech quality. It combines fast and slow compression in order to minimize distortion and to restore the amplitude modulation of speech. VAC is our standard fitting procedure for the hearing aids used in this study. The hearing aid was set to mimic the natural acoustic effect of the pinna by a microphone setting close to omnidirectional.

With activated noise reduction, a fast-acting combination of minimum variance distortion-less response (MVDR) beam-former and a single-channel Wiener post-filter was applied before the VAC. The noise reduction algorithm is based on the finding that a multi-channel Wiener filter can be decomposed into a beam-former and a single-channel Wiener filter, which is better suited for implementation into hearing aids. The noise reduction mainly attenuates interfering sounds originating behind the listener, which should mainly affect the background noise in the current study. To confirm this, the output SNR-improvement of the noise reduction in the hearing aids was measured, which was here defined as the difference between the two frontal talkers and four background talkers. The hearing aid output was measured on a Head and Torso Simulator (HATS). A pair of hearing aids were put on the HATS and the output SNRs of the hearing aids were derived using the phase-inversion technique (Hagerman & Olofsson, 2004). The articulation-index weighted SNR improvements were 6.24 dB and 5.17 dB at +3 dB SNR and +8 dB SNR when noise reduction was on compared to off (see Alickovic et al., 2020).

4.3.1.3 Apparatus

Six loudspeakers at an azimuthal angle of ±22° (competing talkers), ±90° and ±150° (maskers) degrees were set up around the participants. During the task the participant was positioned in the middle of the loudspeakers and the distance to each loudspeaker was 1.2 m. A screen was positioned between the frontal loudspeakers for displaying questions in a way not to cause acoustic shadowing. The setup (**Fig. 4.9**) was inspired by the design of Das et al. (2018). The descriptions for loudspeakers and EEG device are the same as **Section 2.2.2**.

4.3.1.4 Stimuli

86 speech excerpts (i.e., trials) were extracted from publicly available news clips by two different talkers. Half of the trials were spoken by a female talker (T1) and the other half by a male talker (T2). Each segment was 33 seconds long after pauses longer than 200 ms were removed. These news clips were used as target and distractor streams and presented from the frontal loudspeakers. The organization of the male and the female talker being target or distractor as well as their location (left or right loudspeaker) was randomized.

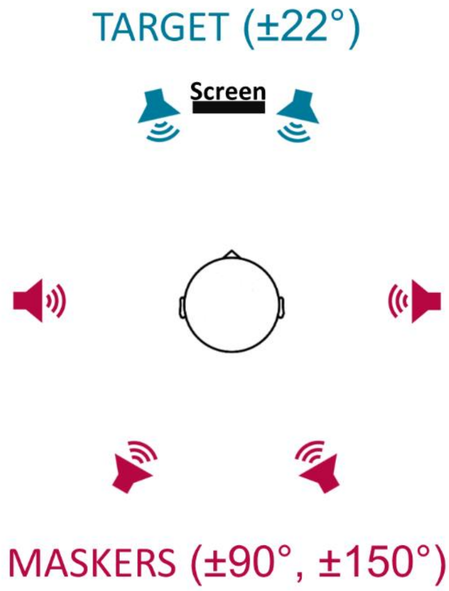


Fig. 4.9 Spatial setup of the task: Six loudspeakers at an azimuthal angle of ±22° (competing talkers – in blue), ±90° and ±150° (maskers – in red) degrees were set up around the participants. The loudspeakers were 1.2 m away from the listener.

As background noise, four-talker babble containing news clip was presented from each loudspeaker. There were total of four loudspeakers in the back, resulting in a 16-talker babble noise. Each 4-talker babble-noise consisted of two female talkers and two male talkers, which were different from the talkers presented at the frontal loudspeakers. The set of four talkers was the same at each background loudspeaker, but it was assured that the identical news clip was not simultaneously presented at two different loudspeakers.

After the two frontal talkers were rms-equalized, the SPL of each of the two frontal talkers was set to 62 dB, respectively. The long-term average spectrum of the babble noise was matched to the long-term average spectrum of the target talkers. The SPL of each 4-talker babble noise was set to 53 dB or 48 dB. Since the summed SPL of the 16-talker babble noise is 6 dB above the SPL of the individual 4-talker babble noise, the SNR of the frontal talkers relative to the background noise was either +3 dB or +8 dB. Note that the effective SNR was even lower, since one of the frontal talkers in the role of a distractor will add to the noise. However, +3 dB and +8 dB will be further used as condition labels for the factor SNR.

4.3.1.5 Procedure

The order of conditions followed a blocked design with the order of blocks being randomized across participants. For each participant, the experiment consisted of four major blocks (20 trials per block). In order to blind the randomization of noise reduction to the participant, the hearing aids were always taken out of the booth and inserted again between the blocks. In two blocks, the noise reduction was switched on and in the other two blocks it was switched off. Within each block, the SNR was fixed at either +3 or +8 dB. The order was randomized such that all four possible combinations of SNRs (+3, +8 dB) and noise reduction (Off, On) occurred once per participant. This randomization was balanced across participants. Over the 20 trials within each block, the target talker and its location changed every fifth trial, such that all four combinations of talker (T1, T2) and position (left, right) occurred within each block in a random succession. To indicate the target talker and its position, before every fifth trial, a 5-s snippet from the talker's voice was presented at the to-be-attended loudspeaker. Before and during each trial, the to-be-attended loudspeaker was also indicated by an arrow at the screen. To acclimatize to the hearing aids, before the start of the experiment, the participants listened to six training trials with noise reduction on.

Participants were instructed to visually fixate a cross in the middle of the screen during listening. The presentation of the sound started with background babble-noise of five seconds. This period mainly served as baseline for the acquired physiological measures. Subsequently, the two news clips were presented at the frontal speakers in the presence of the ongoing background babble-noise (**Fig. 4.10**). After each trial, a statement about the content of the to-be-attended news clip was displayed on the screen, [e.g., "An increasing number of cruise tourists come to Copenhagen." (English translation)]. Participants were asked to indicate whether this statement was correct or wrong. Consequently, the chance level was 50%.

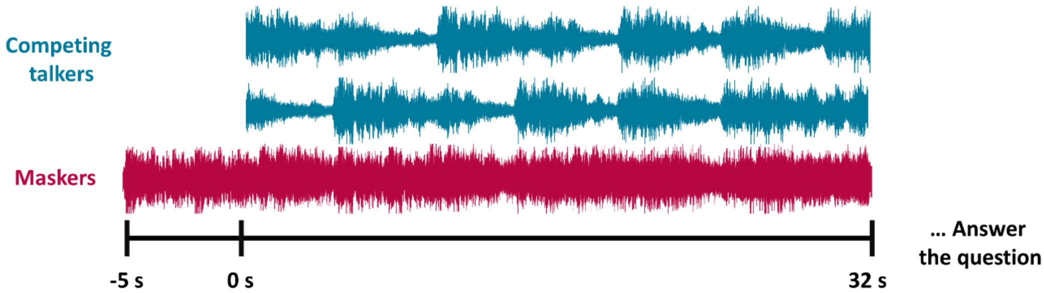


Fig. 4.10 Trial design: Each trial consisted of 37 seconds of 4-talker babble played in the background (in red). The target and distractor speech were presented 5 seconds after the onset of the babble and then continued for 32 seconds (in blue), followed by a yes/no question regarding the content of the attended target audio clip.

4.3.1.6 EEG analysis

The pre-processing of EEG data was done similarly to **Section 2.2.5.1**. Two participants due to poor performance (below average 55%), and three participants due to too few remaining trials to analyze (either due to missing trials or +30% rejected trials) were discarded. For the remaining participants, 3.9% of the channels were detected as bad and interpolated. Also, 5.6% of all trials were rejected with no participant having more than 16.2% of the trials rejected.

Power extraction was also done exactly similarly to **Section 4.2.1.6**. For the phasic power, alpha power ERD in the parietal region was investigated. For the tonic power, theta (exploratory), alpha (hypothesis-driven), and beta (exploratory) power in the parietal and fronto-central regions were analysed. Similar to **Section 4.2.1.6**, theta is defined as 5 - 7 Hz, alpha as 8 - 12 Hz (10 Hz peak in the parietal region; see **Fig. 4.12**), and beta as 13 - 20 Hz.

4.3.1.7 Statistics

LMM was used to investigate the effects of SNR and noise reduction (NR) on performance and EEG power. As most of the participants reported one talker (T1) being more intelligible than the other one (T2), we also considered Talker in our model as well as SNR and NR. Therefore, SNR, NR and Talker were treated as fixed factors and participants were considered as random factors. The MATLAB syntax for LMM implementation was $Dependent \sim 1 + SNR * NR + Talker + (1|Subject\ ID)$, with *Dependent* being either performance or EEG power. In all the models, NR values were coded as -0.5 (Off) and +0.5 (On) and Talkers were coded as -0.5 (T1) and +0.5 (T2). Also, SNR values were centered around 0 (i.e., -2.5 and 2.5 dB) to avoid correlation between the main effects and interactions in the models. The estimates of the LMM (β), t-values and degrees of freedom (t_{DF}) and P values are reported in the Results section.

To analyze changes over time in parietal alpha and correlation between performance and EEG power, similar approaches to **Section 4.2.1.7** were used.

4.3.2 Results

4.3.2.1 Performance

The results of performance showed significant improvement in performance due to increased SNR ($\beta = 0.91$, SE = 0.38, $t_{131} = 2.37$, $p = 0.018$), but there was no such effect for NR ($\beta = 2.79$, SE = 1.91, $t_{131} = 1.45$, $p = 0.147$) nor a significant interaction between the two ($\beta = -0.41$, SE = 0.76, $t_{131} = -0.53$, $p = 0.592$). Interestingly, the strongest improvement on performance was due to the T1 Talker ($\beta = 9.85$, SE = 1.91, $t_{131} = 5.13$, $p < 0.001$). It is worth mentioning that correct response percentages were relatively high for all conditions (above 80%) and the difference between highest and lowest average performances was only 7.35% (**Fig. 4.11**).

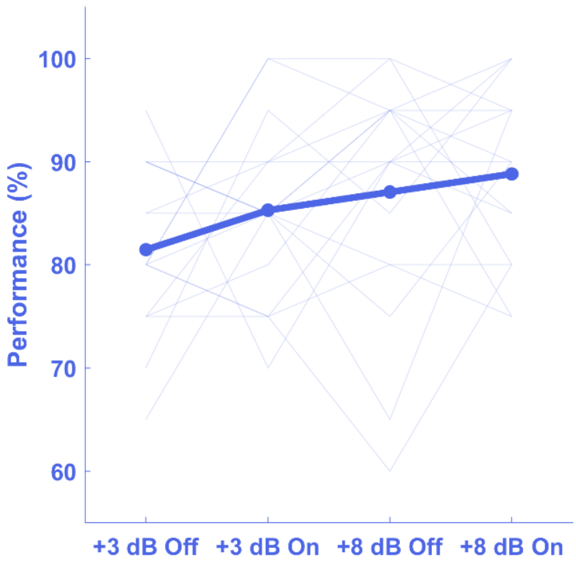


Fig. 4.11 Performance results: The correct percentage of the questions that were answered regarding the contents of the target speech.

4.3.2.2 EEG

4.3.2.2.1 Phasic power

Similar to the first study (**Section 4.2.2.2**), phasic alpha showed desynchronization, but it was not significantly modulated by either of SNR, NR, or Talker in either of the parietal or fronto-cental regions (see **Table 4.4**).

Table 4.4 Results of mixed model based on SNR and NR predictors: estimates of different relative power changes in the parietal region in different bands and phases. Significant p-values are shown in black font.

DF = 131	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = 0.01 SE = 0.46 t = 0.03 p = 0.970	β = 0.54 SE = 0.32 t = 1.70 p = 0.091	**β = 0.98** **SE = 0.47** **t = 2.05** **p = 0.042**	β = 0.45 SE = 0.29 t = 1.54 p = 0.123
NR	β = 1.61 SE = 2.32 t = 0.69 p = 0.486	β = 0.21 SE = 1.60 t = 0.13 p = 0.892	β = 3.81 SE = 2.38 t = 1.59 p = 0.112	β = 2.39 SE = 1.47 t = 1.62 p = 0.107
Talker	β = 0.31 SE = 2.32 t = 0.13 p = 0.891	**β = 3.85** **SE = 1.60** **t = 2.40** **p = 0.017**	**β = 5.43** **SE = 2.38** **t = 2.27** **p = 0.024**	**β = 4.15** **SE = 1.47** **t = 2.81** **p = 0.005**
SNR:NR	β = 0.03 SE = 0.92 t = 0.03 p = 0.973	β = 0.05 SE = 0.64 t = 0.08 p = 0.931	β = 1.49 SE = 0.95 t = 1.56 p = 0.120	β = 0.49 SE = 0.59 t = 0.83 p = 0.407

4.3.2.2.2 Tonic power

Similar to the behavioural data, the Talker factor had the strongest effect, with T1 talker (the more intelligible talker) leading to increased EEG power. This was observed in tonic theta, alpha and beta in parietal (**Table 4.4**) and tonic alpha and beta in the fronto-central regions (**Table 4.5**). Higher SNR only led to higher tonic alpha in the parietal region (**Fig. 4.12**) and NR did not change any of the bands in any of the explored locations.

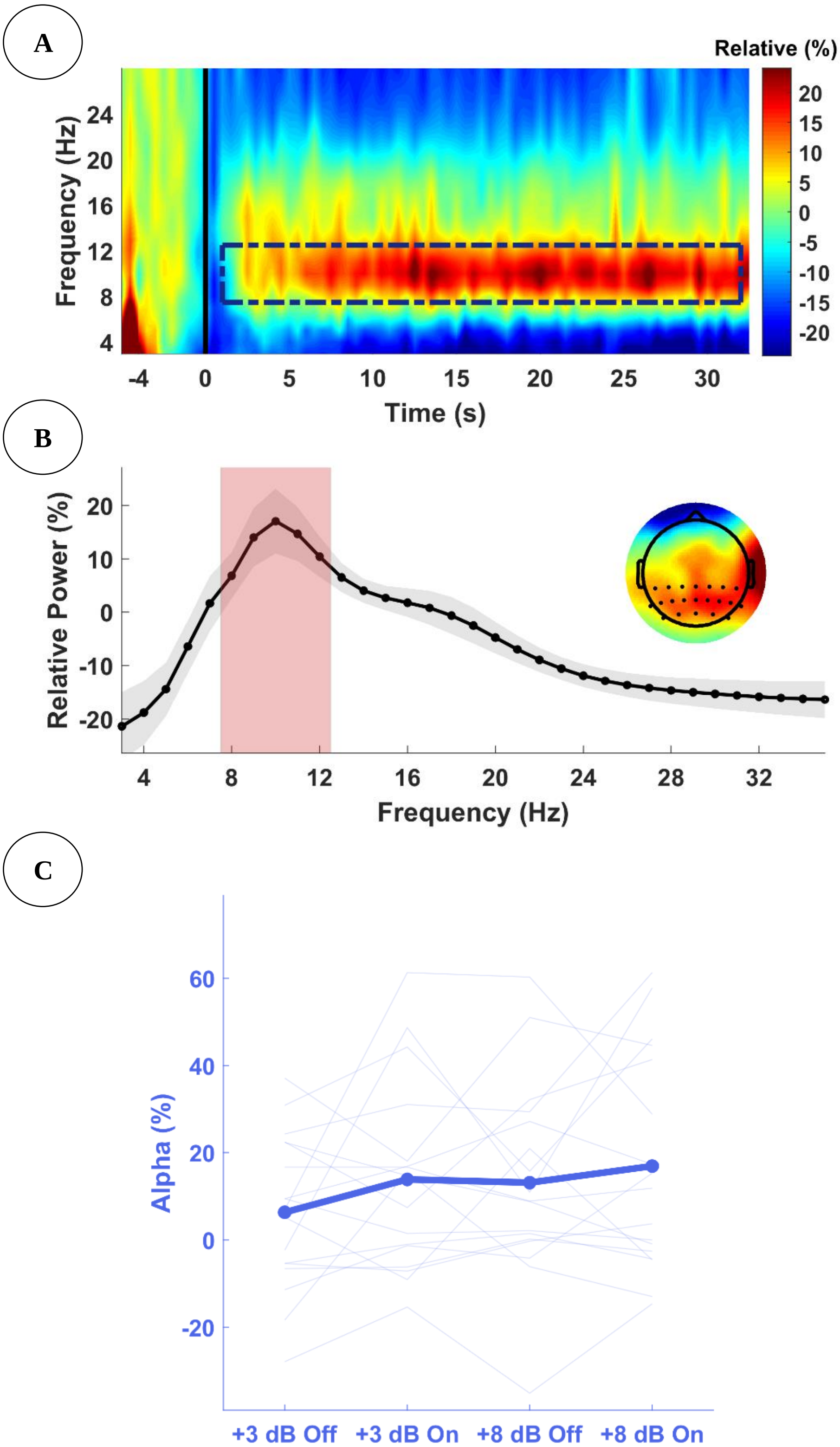


Fig. 4.12 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR in the highlighted window of panel A which showed significant effect.

4.3.2.2.3 Changes over time

The analysis of changes over time showed a significant positive slope of parietal alpha power in grand averaged data. Investigation of each individual condition showed significant slopes for all except the condition hypothesized to be hardest (+3 dB SNR, NR Off). The analysis of 5-s slope across the trial also showed a significant positive slope only in the first five seconds for all the conditions except +3 dB NR off, after FDR correction (**Fig. 4.15**).

Table 4.5 Results of mixed model based on SNR and NR predictors: Estimates of different relative power changes in the fronto-central region in different bands and phases. Significant p-values are shown in black font.

DF = 131	Phasic	Tonic		
Band Predictor	Alpha	Theta	Alpha	Beta
SNR	β = -1.05 SE = 0.73 t = -1.43 p = 0.152	β = 0.46 SE = 0.31 t = 1.49 p = 0.137	β = 0.44 SE = 0.46 t = 0.95 p = 0.342	β = 0.25 SE = 0.29 t = 0.88 p = 0.377
NR	β = -2.11 SE = 3.65 t = -0.57 p = 0.564	β = 0.08 SE = 1.57 t = 0.05 p = 0.958	β = 2.59 SE = 2.34 t = 1.10 p = 0.271	β = 2.17 SE = 1.45 t = 1.49 p = 0.138
Talker	β = -2.43 SE = 3.65 t = -0.66 p = 0.507	β = 1.29 SE = 1.56 t = 0.82 p = 0.410	**β = 4.92** **SE = 2.34** **t = 2.09** **p = 0.037**	**β = 3.60** **SE = 1.45** **t = 2.47** **p = 0.014**
SNR:NR	β = -0.61 SE = 1.46 t = -0.42 p = 0.674	β = 0.73 SE = 0.62 t = 1.17 p = 0.244	β = 1.18 SE = 0.93 t = 1.25 p = 0.210	β = 0.76 SE = 0.58 t = 1.31 p = 0.189

4.3.2.3 Correlation

The results for Pearson skipped correlation between performance and EEG power in different electrode regions are shown in **Table 4.6**. Based on the 95% CI the only significant correlations occurred in tonic theta, both in parietal and fronto-central regions (**Fig. 4.16**). That is, increase in theta power was co-modulated with increase in performance accuracy.

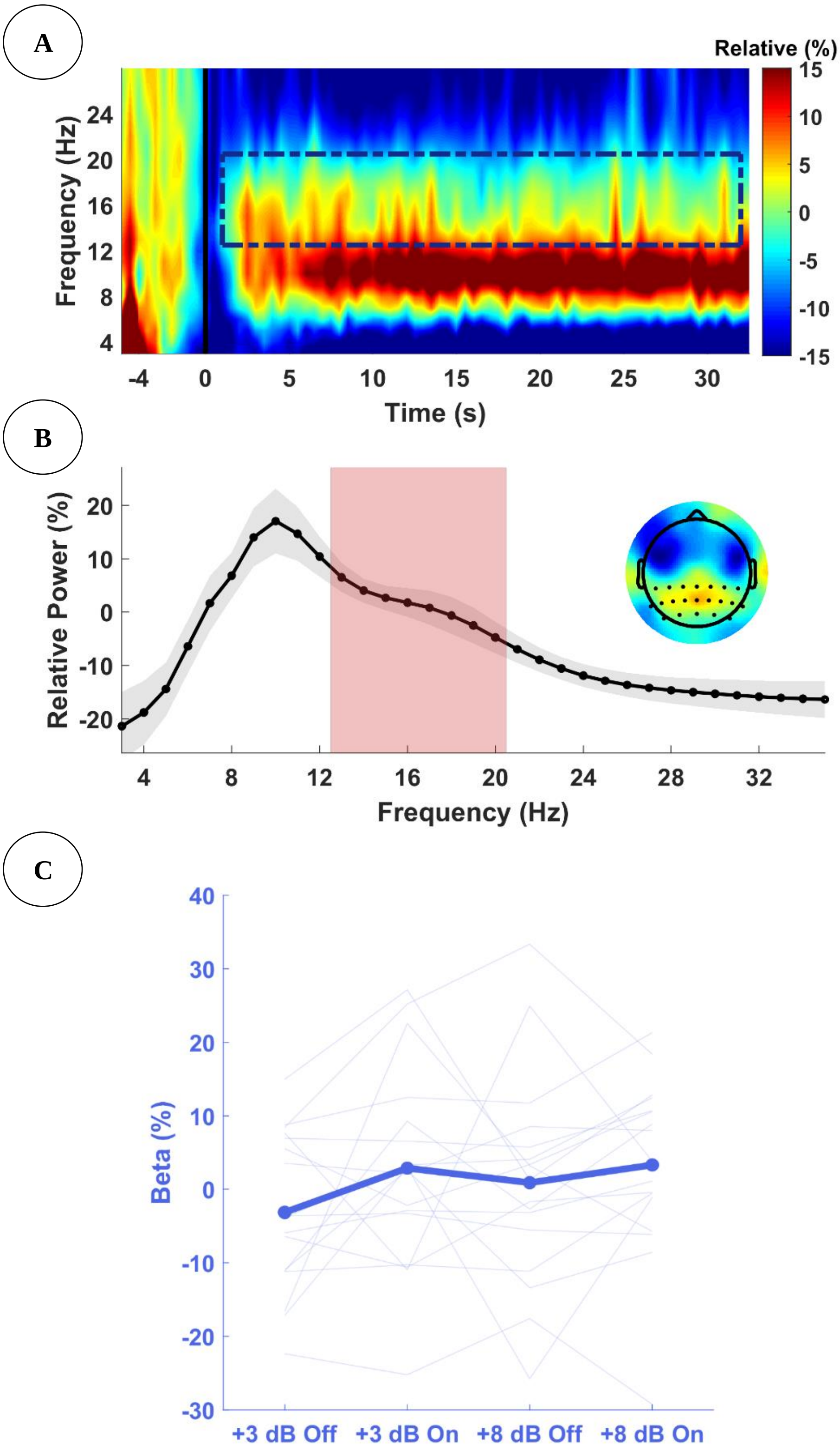


Fig. 4.13 Power changes over time: A) Grand average spectrogram , B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR in the highlighted window of panel A which showed significant effect.

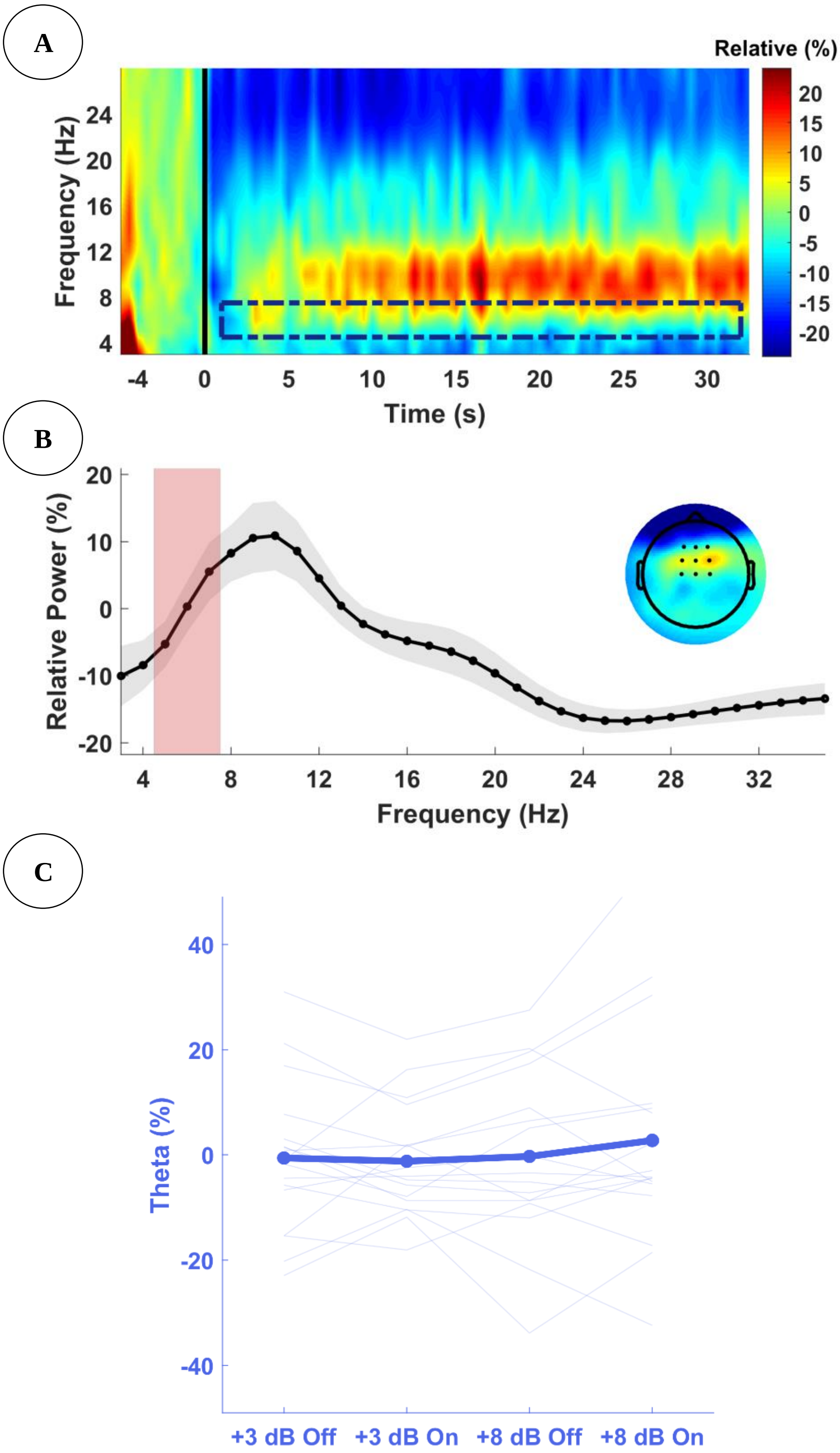


Fig. 4.14 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions. C) The modulation of beta power by SNR in the highlighted window of panel A which showed significant effect.

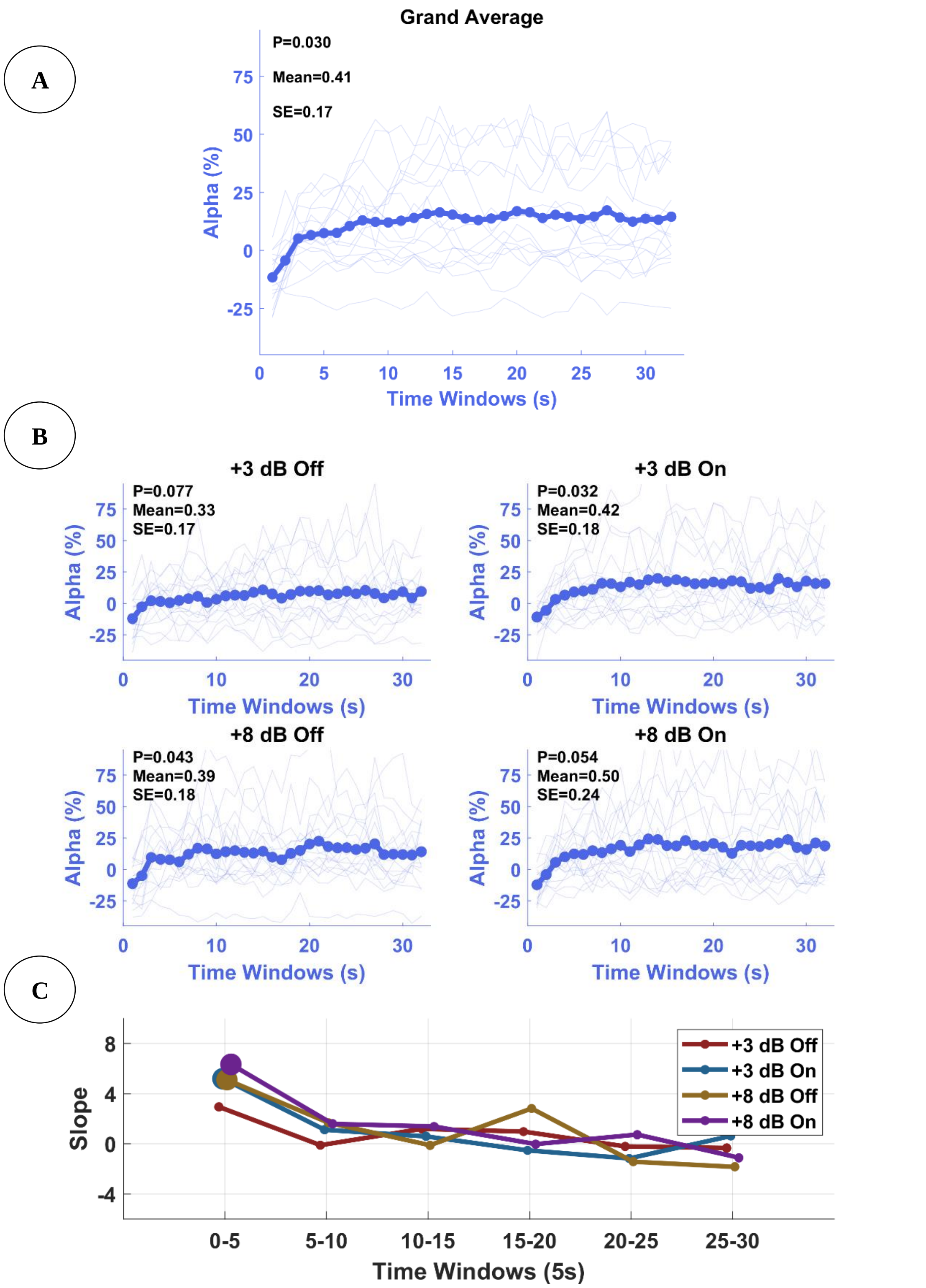


Fig. 4.15 Changes of parietal alpha power over time for A) grand average data and B) each condition in 1-sec windows. C) The analysis of slope within 5-s intervals showed significant positive slopes in the first five seconds (shown in large dots) for all except the hardest hypothesize condition (i.e., 3 dB SNR, NR Off).

Table 4.6 Pearson skipped correlation between performance and EEG power in different regions, bands, and phases. Significant correlations are shown in black.

	Phasic	Tonic		
Band / Electrodes	Alpha	Theta	Alpha	Beta
Parietal	r = 0.05 CI = [-0.15 0.26]	**r = 0.27 CI = [0.06 0.45]**	r = 0.14 CI = [-0.09 0.34]	r = 0.07 CI = [-0.17 0.31]
Fronto-central	r = -0.17 CI = [-0.38 0.07]	**r = 0.32 CI = [0.12 0.48]**	r = 0.10 CI = [-0.20 0.35]	r = 0.26 CI = [-0.01 0.48]

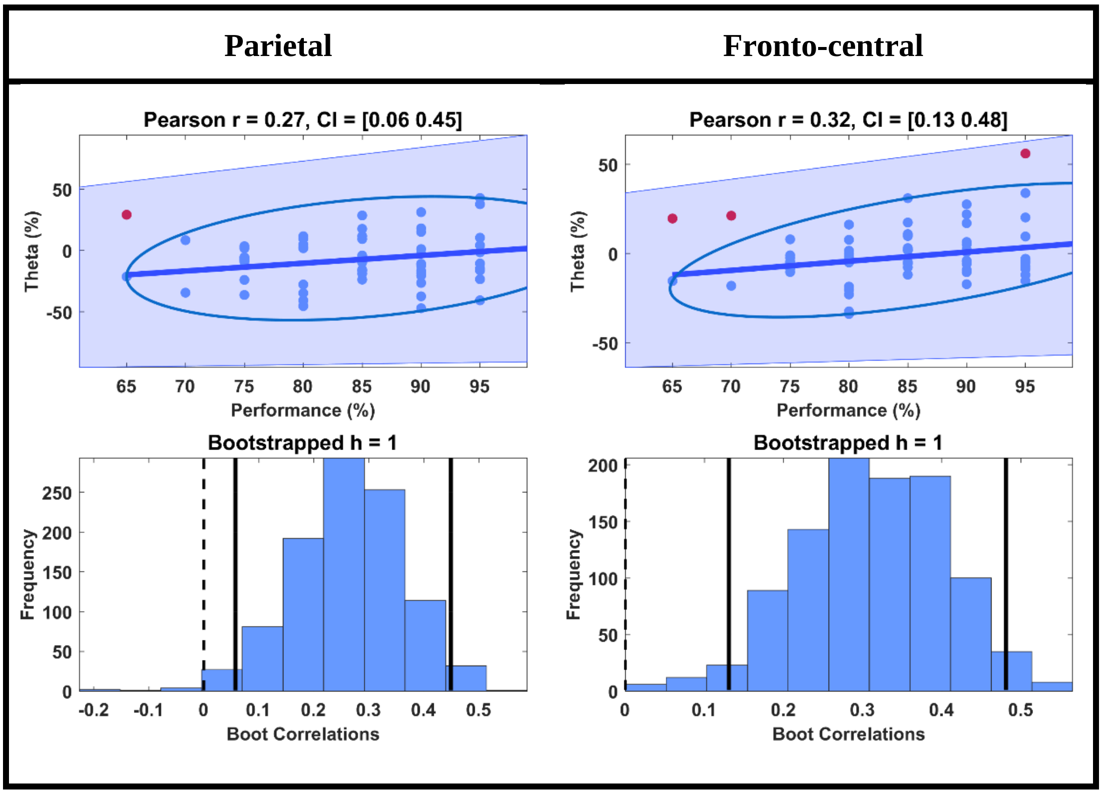


Fig. 4.16 Pearson's skipped correlation with bootstrapping between performance and tonic theta power in the parietal and fronto-central regions. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI. Both showed statistically significant correlation (h = 1) with performance.

4.3.3 Summary

The second study showed that alpha power in hearing-impaired participants were increased due to higher SNR and change of Talker in a continuous-speech paradigm. The results of this study were in line with the first study that alpha power decreases in continuous speech under demanding conditions. However unexpectedly, the strongest effect of the second experiment was due to Talker. Both performance and EEG power were significantly affected by Talker. The more intelligible talker had the highest impact on improving performance and led to increase in tonic theta, alpha, beta in the parietal and tonic alpha and beta in the fronto-central region.

4.4 Discussion

4.4.1 Overview

In the two experiments of this chapter, it was investigated that how effortful listening manifests in more demanding situations while listening to long, uninterrupted speech in noise. In the first experiment, manipulation of SNR showed that alpha power in the parietal lobe was lower in the harder condition (low SNR) compared to the easy condition (high SNR). In the second experiment, manipulation of SNR showed a similar pattern (i.e., less parietal alpha for more demanding conditions) without significant effects of hearing aid noise reduction. However, an exploratory finding in the second experiment showed that the talker influenced the listeners (in terms of performance and expenditure of effort) more than changes of SNR and noise reduction scheme.

4.4.2 Alpha in demanding continuous discourse

Using alpha power as an outcome measure for listening effort has resulted in contradictory results in previous studies. While most of the studies that have used short-stimulus paradigm suggest the relationship between alpha power and task difficulty is direct, i.e. more difficulty equals increased alpha (e.g., Obleser et al., 2012; Petersen et al., 2015; Wisniewski, Thompson, et al., 2017; Wöstmann et al., 2015, 2017), studies with longer stimuli have shown the inverse, i.e. more difficulty equals decreased alpha (Hjortkjaer et al., 2018; McMahon et al., 2016; Miles et al., 2017).

Therefore, it is possible that alpha power during short and continuous speech reflect different cognitive aspects.

As mentioned in **Chapter 1**, two common and conflicting theories on alpha power in effortful situations exist. One theory explains that increased alpha is a sign of suppression of unattended sound sources (Holm et al., 2009) and inhibition of task-irrelevant cortical regions (Klimesch, 2012), which as a consequence should increase alpha with increased difficulty. On the other hand, the "cortical idling" theory states that synchronized (i.e. increased) alpha is a correlate of a deactivated cortical network (Pfurtscheller, 2001) and therefore listening effort should lead to a desynchronized state of alpha power.

Our results cannot be fully explained by either of the two theories. The inhibition theory indicates that more demanding situations should lead to increased alpha power, whereas in the current study it was the opposite. "Cortical idling" also cannot explain why there is a distinct synchronization of alpha power during attending to a speech which requires complex working memory processing.

One possible explanation for our results comes from identifying commonalities with previous literature that have also found alpha power to decrease in more demanding situations. Jensen et al., 2002, speculated the reason for different patterns of alpha power with workload. In their study, participants performed the Stenberg task to see how parietal alpha alters with higher workloads. They observed increased alpha power with increased workload which was opposite to another working memory study (n-back task), in which decreased alpha activity was observed with higher demand (Gevins et al., 1997). They concluded that in the Stenberg task the brain response is different when the encoding and retention phases are temporally independent from each other, compared to an n-back task where these phases are overlapping and require a constant update of information in the working memory. Given the nature of the stimuli in the current study, sustained attention and constant updating of working memory is required over 30 seconds of speech presentation. The entangled encoding and retention phases might call for decreased alpha activation when it is more difficult. This notion goes along with other studies that showed optimal sustained attention performance is linked to greater alpha oscillation (Dockree et al., 2007; Hjortkjaer et al., 2018).

In the design of two studies in this chapter, a narrow range of SNRs was chosen to mitigate the effects of stimulus-related responses. However, it is plausible that observed alpha changes is due to tracking of acoustic features of the speech (i.e., stimulus-related response). In that sense, it might be that not just alpha power, but also other bands in brain signals experience an overall increase while listening to more intelligible speech. The evidence for this suggestion came when listening to T1 talker (the more intelligible talker) in the second study, which led to increase of theta, alpha and beta power in the brain. It is also important to note that performance was also better with the T1 talker and subjective reports from participants also indicated that listening to T1 talker was actually easier. Therefore, it is also possible that alpha power reflected performance of the participants. This issue will be further explored in **Chapter 5**.

In sum, if alpha oscillations in continuous speech is indeed inhibiting task-irrelevant regions of the brain, then higher task demands should have increased alpha power. Therefore, it is possible that in this continuous paradigm, alpha (or other bands) is not reflecting effort and might be just tracking the acoustic features of the speech.

4.4.3 Changes of alpha power over time

One of the advantages of using long speech (in this case around 30 s) was that it allowed for studying how alpha power modulates over time. Based on the overall time-frequency results, alpha power was more prominent towards the end of the trial compared to the beginning. For this reason, I decided to look at the changes of alpha power over time using first-degree polynomial curve fitting.

In the first experiment, there was no positive slope throughout the trials in grand average data nor in any of the 5-s windows within a trial. This lack of significance might have happened due to low number of participants that led to type II error rate.

However, in the second experiment with a larger population, the slope of alpha power was significantly positive in the grand average data (i.e., alpha power increased over time during trials). To further narrow the time dynamic of changes for this positive slope, shorter 5-s window periods were investigated. Interestingly, the only significant positive slope happened within the first 5 seconds of the trial for all the conditions except the hypothesized-hardest condition (+3 dB SNR, NR Off). As observed in **Fig. 4.12**, the alpha power after the stimuli begins with a desynchronized phase and then shifts to synchronized phase. Therefore, the slope of the first 5-s window can be interpreted as the magnitude of change in transition of alpha power from the desynchronized to the synchronized phase. It might be that if this transition from baseline is stronger in one condition (i.e., the slope is more positive), tonic alpha power would also be set to be higher for the rest of the listening for over 30 seconds. Therefore, looking at the changes of alpha power in the first few seconds of such a long speech might be enough to discriminate the difficulty of the ongoing listening task.

4.4.4 Correlation of EEG power and performance

As mentioned in **Chapter 1**, there is a considerable number of studies that have shown that EEG power (more specifically alpha power) can be correlated to behavioural performance. In the first study, no significant correlation between EEG and performance was observed, which might be another indication of small population leading to type II error rate.

In the second study, there was a correlation between theta power and performance (in the parietal and fronto-central regions). Such a correlation may be related to speech tracking with EEG signals using different methods such as stimulus reconstruction (e.g., Alickovic et al., 2020; Das et al., 2018; O'Sullivan et al., 2015; Petersen et al., 2017). In fact, using stimulus reconstruction, it has been shown in the very same dataset (second experiment) that higher SNR and activation of the noise reduction scheme enhanced the neural representation of speech and reduced the neural representation of background noise during such a long stimuli (Alickovic et al., 2020). As stimulus reconstruction mainly relies on the theta band (4-8 Hz) of the EEG signal, it is possible that this mediates the correlation.

4.4.5 Limitations

In the first experiment, one limitation is the low number of participants recruited (n = 8). Due to small number of participants, chances for type II error rate in the statistical analyses were high and it might be one explanation that the results lacked the significant findings of the second experiment.

In the second experiment, an unexpected strong effect was observed on both behavioural and EEG power data due to differences between the two talkers used as stimuli. We cannot infer what caused this difference. Even though the stimuli were normalized, the subjectively reported difference in intelligibility could have been caused by a difference in energetic masking due to spectral differences. This would lead to increased effort at a comparably early stage of the neural encoding of the speech signal. As stated earlier we had two different talkers, each spoken by a different gender (T1 by a female vs. T2 by a male) which led to higher speech intelligibility in the female talker. While we did not infer the effect of gender from the effect of the talker, since we only presented one example of each gender, there have been other studies that have shown female talkers have led to higher speech intelligibility (Kwon, 2010; Yoho et al., 2019). With our design, we could not provide further insight into differences caused by gender, but we could show that both performance and EEG power were affected after participants consistently expressed intelligibility concerns.

One other limitation of the second study is in how far the hearing aid fitting strategy generalizes to recommended fitting strategies. Here we used non-individualized, closed tips for better control of the signal that enters the auditory pathway. However, in practice, patients with mild to moderate hearing loss (and intact hearing in the low frequency range) would have been fitted with open domes. Therefore, noise reduction would only affect higher frequencies, such that the overall effect of noise reduction might be reduced for patients with open fit. Future studies should address this to claim higher ecological validity. Furthermore, the effect of noise reduction was only quantified once on a dummy head. In future studies, in-situ measurements should be conducted to quantify the effect of noise reduction in the individual patient.

4.5 Conclusion

In two experiments EEG signals were used to assess aspects of listening effort of hearing-aid users in a continuous speech setting, presented from either a right or left target in the presence of noise. In the first experiment, SNR, and in the second experiment, SNR and noise reduction schemes were manipulated for hearing aid users. Both experiments showed that more demanding conditions led to less activation of tonic alpha in the parietal region. In the first experiment this was due to SNR, and in the second experiment due to SNR and talker intelligibility. While the initial hypothesis was that more demanding listening conditions should lead to increased parietal alpha, these results suggest there the pattern of alpha power in long speech is the opposite and it may not reflect listening effort.

Chapter 5
Ear-EEG as a wearable technology to measure listening effort

The study in this chapter was a joint project with UNEEG medical A/S, Denmark. TSA's contributions were data analysis and interpretation.

5.1 Introduction

5.1.1 Ear-EEG

So far, this thesis has been focusing on scalp EEG. Many studies within auditory and/or cognitive neuroscience, have used EEG due to its non-invasiveness and excellent temporal resolution, along with several other advantages (Luck, 2014). However, despite high quality recordings by EEG, it is not yet suitable as a wearable technology and can be hardly used in real-life applications (Looney et al., 2012), which makes it limited mostly to the laboratory experiments.

In the spirit of introducing a new wearable technology that can be used as an ambulatory measurement of brain electrical activities, EEG signals now can be picked up by ear-EEG electrodes (Bleichner et al., 2015; Looney et al., 2012). Despite the attraction of ear-EEG as a wearable technology, it still suffers from poorer spatial resolution compared to scalp EEG, as the signals are only limited to the brain areas close to the ears (Kappel et al., 2019). Nonetheless, researchers have shown that traditional analyses, such as P300 detection (Bleichner et al., 2015), frequency-domain representation (Mikkelsen et al., 2015), or steady-state evoked potential (Ahn et al., 2018), commonly studied by scalp EEG are also relevant in ear-EEG. However, listening effort has not yet been explored using ear-EEG.

The first generation of ear-EEGs required conductive gel to be applied between the electrodes and skin (e.g., Kidmose et al., (2013), Fiedler et al., (2017)), and thus they are referred to as wet ear-EEG. Recent improvement in hardware design of ear-EEGs have removed the need for conductive gel and hence they are called dry ear-EEG. Such hardware improvement made ear-EEGs even more valuable as a potential wearable technology, as it is now more user-friendly, comfortable and would enable the user to insert the device without assistance (Kappel et al., 2019).

Both wet and dry ear-EEGs have more advantages as well. The tight fit between the earpiece and the ear helps the electrodes to be held firmly in place and this can reduce motion or muscular movement artifact. Placing the electrodes next to the ear canal which is a strong electrically conductive medium due to earwax diminishes the interference from external electrical fields. These are some advantages that the signals picked up by ear-EEG can benefit compared to when they are picked up by scalp EEG (Looney et al., 2012).

5.1.2 Objectives

The study in this chapter, similar to those in **Chapter 4**, investigated the changes in brain activities during long, uniteruppted speech-in-noise task. The only variable for this study was SNR (-16, -8, -4, +8 dB) to manipulate task demand on normal-hearing participants. The main distinction between this study and those mentioned in **Chapter 4**, was to implement ear-EEG as well as scalp EEG to look for changes in listening effort. In **Chapter 4**, the results showed that the power in scalp EEG can be modulated by manipulation of task demand, therefore, the aim of this chapter is to compare the changes in ear-EEG to its scalp EEG counterpart.

5.2 Study design

5.2.1 Participants

Fifteen (4 females) normal-hearing Danish-speaking adults (average age of 42.4 ± 11.4 years) participated in this study. One additional participant was excluded due to a problem in sending triggers to the EEG device. All participants signed a written consent form before the experiment. Ethical approval of this study was obtained from the Research Ethics Committees of the Capital Region of Denmark. No participant suffered from neurological or hearing disorders. The pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz (PTA4) were tested for hearing abilities and confirmed to be below 25 SPL HL.

5.2.2 Apparatus

The experimental setup consisted of five loudspeakers positioned around the participant at 1.2 m distance. The target loudspeaker was positioned 0° azimuth in front of the listener. The background noise, consisting of 4-talker babble noise, was presented from four loudspeakers (16-talker babble in total) located at ±90° and ±150° azimuth. The spatial setup of the test is illustrated in **Fig. 5.1**.

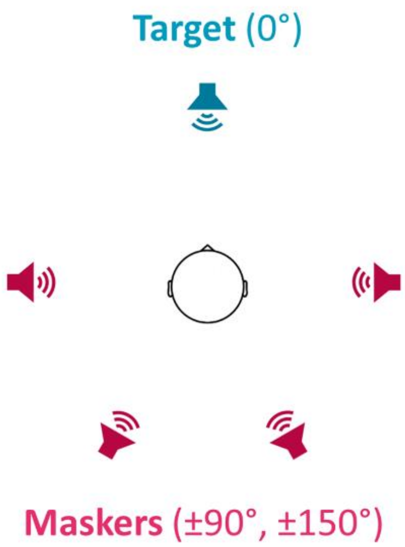


Fig. 5.1 Spatial setup of the task; Five loudspeakers positioned around the participant at 1.2 m distance. The target loudspeaker (in blue) was positioned 0° azimuth in front of the listener. The background noise, consisting of 4-talker babble noise (in red), was presented from four loudspeakers located at ±90° and ±150°.

The descriptions for loudspeakers and scalp EEG device are the same as **Section 2.2.2**. The ear-EEG recordings were acquired with a sampling rate of 1000 Hz by a 32-channel portable TMSi MOBITA EEG amplifier (TMSi, Netherlands). In addition, the amplifier enabled active shielding (guarding) of the ear-electrodes all the way to the backside of each of the 12 electrodes. For each participant, earmould impressions were acquired in a session before the test in order to make personalized ear-EEG.

5.2.3 Stimuli

Non-dramatic Danish news clips of neutral contents were used for the target speech (33 s) and16-talker babble noise (38 s). The A-weighted SPL of the babble was fixed at 70 dB overall (64 dB each). The level of the target was varied across trials from 54-78 dB to generate four different SNRs: -16, -8, -4 and +8 dB. In this study, SNR was defined as the long-term average sound level of the target signal (with pauses longer than 200 ms being cut out) compared to the background noise.

5.2.4 Procedure

There were 21 trials for each SNR, randomly distributed across 84 trials with an additional 4 training trials in the beginning of the test. Each trial (**Fig. 5.2**) consisted of 38 seconds of 16-talker babble played in the background. The target was presented 5 seconds after the onset of the babble (i.e., after the baseline period) and then continued for 33 seconds, followed by a two-choice question regarding the content of the attended target audio clip. Participants were given a rest period every 28 trials. The percentage of correct answers were considered as performance accuracy.

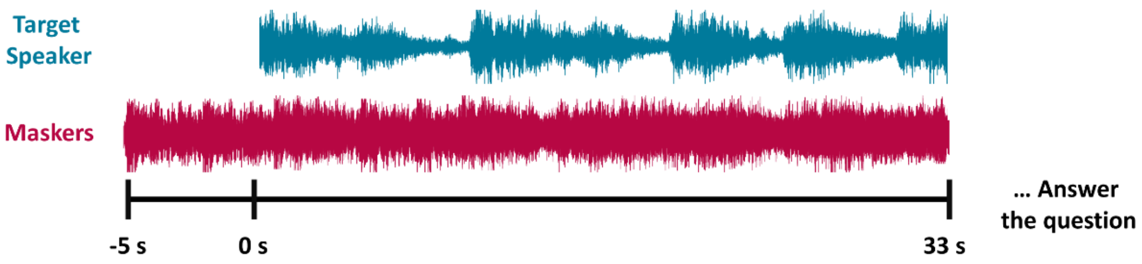


Fig. 5.2 Trial design; Each trial consisted of 38 seconds of 16-talker babble played in the background (in red) with 33 seconds of target speech starting 5 s after the onset of background noise (in blue). A two-choice question regarding the contents of the attended target audio clip was asked from the participants.

5.2.5 Scalp EEG analysis

The pre-processing of EEG data was done similar to **Section 2.2.5.1**. In total, 1.8% of the channels were detected as bad and were interpolated. Also, 9.2% of all trials in this experiment were rejected. No participant had more than 23.8% of trials rejected.

For the resulting signals two different referencing methods were applied. The first method was the common average referencing which is acquired by subtracting the average of all channels from any single channel (Luck, 2014). The second method was the symmetrical bipolar referencing which was achieved by subtracting any two channels mirrored to each other in opposite hemispheres (e.g., T7 - T8). The latter referencing approach allowed us for more direct comparison between the results of scalp EEG and ear-EEG (see **Section 5.2.6**). The resulting signals were used for power extraction which was also done similar to **Section 4.2.1.6**.

Based on the spectrum of averaged data using common referencing (see **Fig. 5.5**) there was a peak at 10 Hz in parietal and using symmetrical referencing (see **Fig. 5.8**) there was a peak at 12 Hz. Therefore, we define the range of alpha power as 8-14 Hz in this experiment to contain both peaks and have a similar range for both common and symmetrical referencing. An exploratory analysis of tonic theta (5 – 7 Hz) and tonic beta power (15 – 20 Hz) was undertaken.

5.2.6 Ear-EEG analysis

For the analysis of the ear-EEG, first power line noise was rejected with a 50-Hz notch filter with a quality factor of 25. Then a 3rd-order zero-phase Butterworth bandpass filter with cut-off frequencies of 1-40 Hz was applied to the data and the resulting signals were down-sampled to 256 Hz. In total, 5 trials were missed in the ear-EEG data due to trigger issues.

The first step in using ear-EEG was to select the best pair of electrodes, one in each ear. Given that the quality of signals in each electrode varied from participant to participant, we decided to choose the most consistent electrodes based on their normalized standard deviation (SD). For this purpose, the SD of all the electrodes within each participant were normalized to the smallest SD in that participant (except when there was no connection for an electrode). Any value further from 1 showed larger signal variance and most probably contained noise. **Fig. 5.3** summarizes the normalized SD across all electrodes and participants (in both left and right ears). As shown, K electrode had the most consistent normalized SD and thus it was chosen as the main electrode for further analysis. However, in two participants the quality for K electrode was quite poor and instead, F electrode was used in order to replace it. The location of F electrode is also very close to K electrode, as both are placed close to ear canal (**Fig. 5.3** bottom panel).

After finding the best pair in each participant, the two electrodes were subtracted from each other (similar to symmetrical referencing in the scalp EEG). The rationale behind the choice of referencing in ear-EEG was that 1) independency from any external referencing electrode while using ear-EEGs and 2) subtracting channels between ears would result in a bigger amplitude signal (compared to referencing to another channel in the same ear which would result in a smaller amplitude signal and thus more susceptible to noise). After that, bad trials were rejected by visual inspection. In total, 10.8% of all trials in the ear-EEG data were rejected while no participant had more than 27.2% of trials rejected.

Power extraction on resulted signals was done similar to power extraction in scalp EEG. Based on the spectrum of the ear-EEG signals there was a peak at 12 Hz. Therefore, we explored phasic and tonic alpha power (8 – 14 Hz) (see **Fig. 5.11**). An exploratory analysis of tonic theta (5 – 7 Hz) and tonic beta power (15 – 20 Hz) was undertaken.

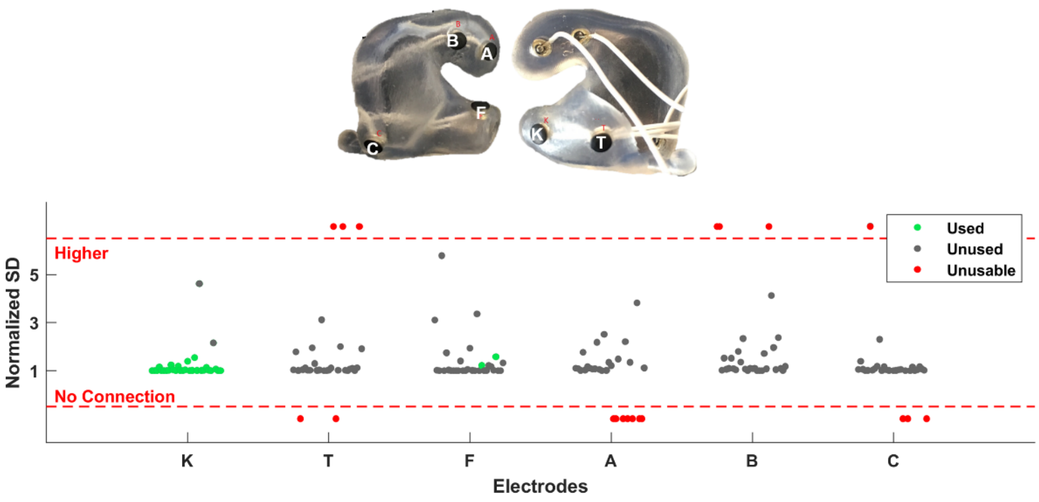


Fig. 5.3 Top row shows the location of the dry ear-EEG electrodes, consisted of 6 electrodes in each ear: A, B, C, F, K, T. Bottom row shows the normalized standard deviation (SD) for each electrode in ear-EEG for all participants' both ears (180 in total). The green channels (30) are the electrodes that are used for further analysis (28 Ks and 2 Fs). The unusable channels below the "*No Connection*" or above "*> 6 x Normalized SD*" had no output or were too noisy, respectively.

5.2.7 Statistics

For both scalp EEG and ear-EEG, LMM was used to estimate the effects of SNR on performance and EEG power in different bands. SNR was the fixed factor, and the participants were the random factors. The MATLAB syntax to implement LMM was $Dependent \sim 1 + SNR^3 + (1|Subject\ ID)$ with *Dependent* variable being either performance or EEG power. *SNR* values were centered around 0 (i.e., -11, -3, +1, +13 dB) in the model to avoid correlations between the linear, quadratic, and cubic effects. The degree of the model was chosen based on the sigmoid pattern of grand average results of both performance and alpha power (see **Section 5.3**) that were used in this model.

Due to long length of the stimuli, changes of EEG power over time were also investigated. The procedure was similar to the description in **Section 4.2.1.7**. For correlation analysis between EEG power and performance, Pearson skipped correlation was used as described in **Section 2.2.7**.

5.3 Results

5.3.1 Performance

The performance was significantly improved with increasing SNR ($\beta = 6.78$, $SE = 0.96$, $t_{56} = 7.03$, $p < 0.001$) without any quadratic effects of SNR ($\beta = 0.01$, $SE = 0.02$, $t_{56} = 0.37$, $p = 0.710$) and significant cubic effects of SNR ($\beta = -0.03$, $SE = 0.01$, $t_{56} = -5.42$, $p < 0.001$). The average results for each condition (**Fig. 5.4**) show that for the two most difficult conditions (-16 and -8 dB) the performance was just slightly above chance level (50%).

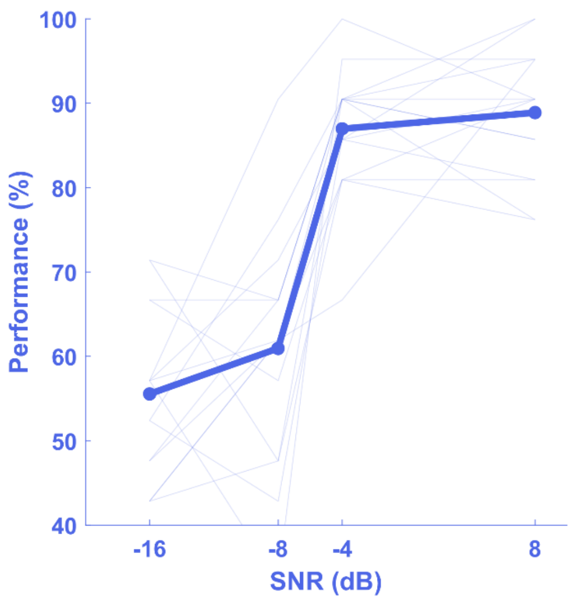


Fig. 5.4 Performance correct percentage based on the two-choice questions regarding the contents of the attended speech.

5.3.2 Scalp EEG – Common

Using common referencing in parietal region (**Table 5.1**), there was no significant change of phasic alpha due to SNR. However, for tonic power, theta, alpha and beta were modulated by SNR and only theta and beta were modulated by SNR^3. The changes in tonic alpha are shown in **Fig. 5.5**.

The analysis of changes over time showed significant slope in alpha power over 32 seconds in grand averaged data (**Fig. 5.6 top panel**). Investigating each condition separately, it was observed the two most difficult conditions (-16 and -8 dB) were the only conditions with significant positive slope during 32-second period (**Fig. 5.6 middle panel**). Dividing the alpha power into 5-s windows, the only significant positive slope happened during the first 5 second in the two easiest conditions (-4 and +8 dB) (**Fig. 5.6 bottom panel**).

The results of skipped Pearson with common referencing in parietal region (**Table 5.2**) showed significant correlation between performance and tonic theta, alpha and beta bands. That means that all the bands which were significantly modulated by SNR also showed significant correlation with performance. The correlation between tonic alpha and performance is shown in **Fig. 5.7**.

Table 5.1 Results of mixed model based on SNR predictor: estimates of different relative power changes using common referencing in the parietal region in different bands and phases. Significant p-values are shown in black font.

DF = 57	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = -0.26 SE = 0.91 t = -0.29 p = 0.772	**β = 2.52 SE = 0.74 t = 3.36 p = 0.001**	**β = 2.68 SE = 0.96 t = 4.48 p < 0.001**	**β = 2.33 SE = 0.59 t = 3.89 p < 0.001**
SNR^2	β = 0.01 SE = 0.02 t = 0.32 p = 0.746	β = -0.03 SE = 0.01 t = -1.87 p = 0.066	β = -0.41 SE = 0.02 t = -1.98 p = 0.052	β = -0.01 SE = 0.01 t = -0.78 p = 0.434
SNR^3	β = 0.001 SE = 0.006 t = 0.19 p = 0.842	**β = -0.01 SE = 0.005 t = -2.30 p = 0.025**	β = -0.01 SE = 0.006 t = -1.91 p = 0.061	**β = -0.01 SE = 0.004 t = -2.74 p = 0.008**

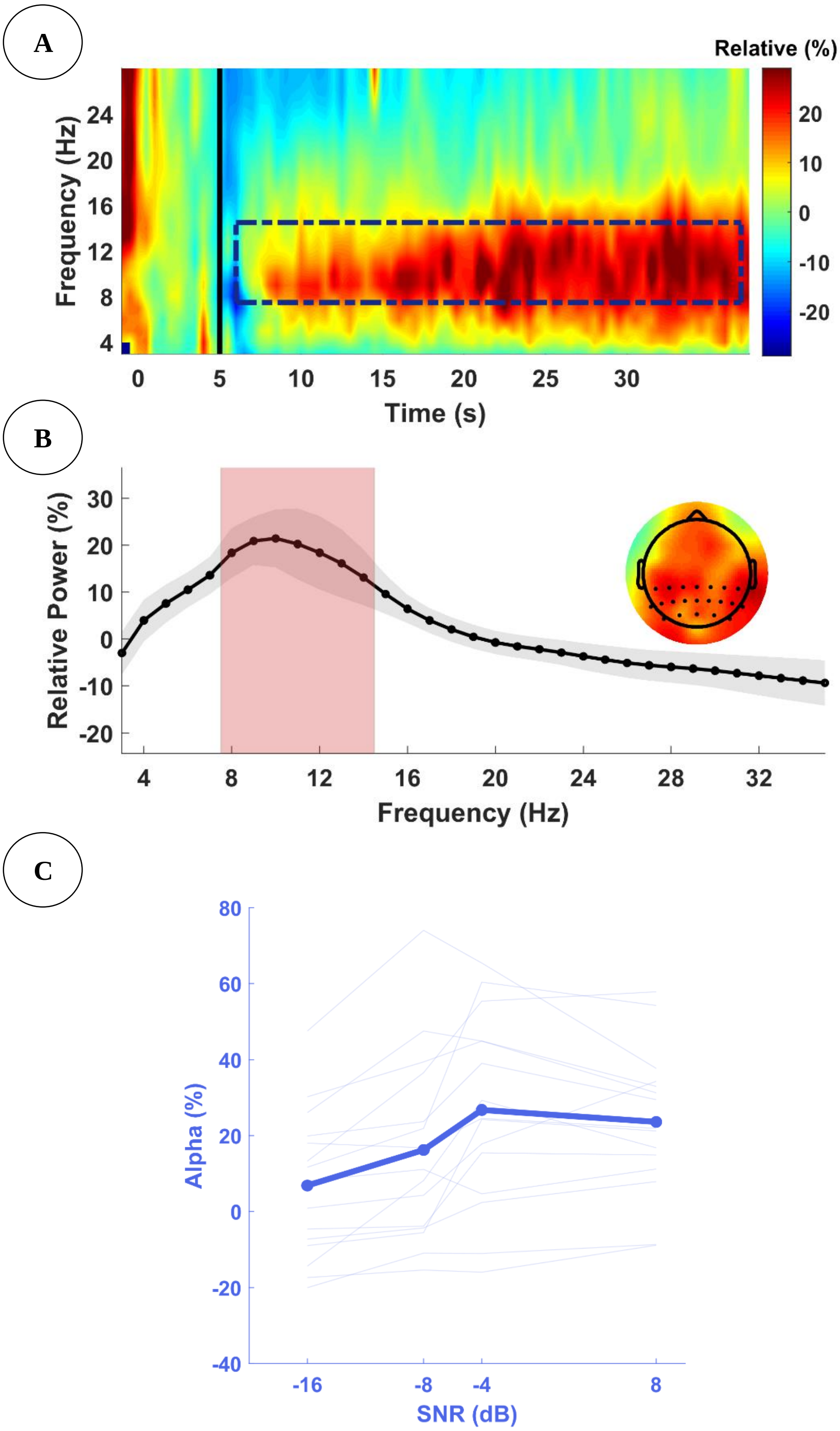


Fig. 5.5 Power changes over time: A) Grand average spectrogram (, spectrum and topographic map in the highlighted time window of panel A of all participants and conditions after common referencing. C) The modulation of alpha power by SNR in the highlighted window of panel A which showed significant linear effect.

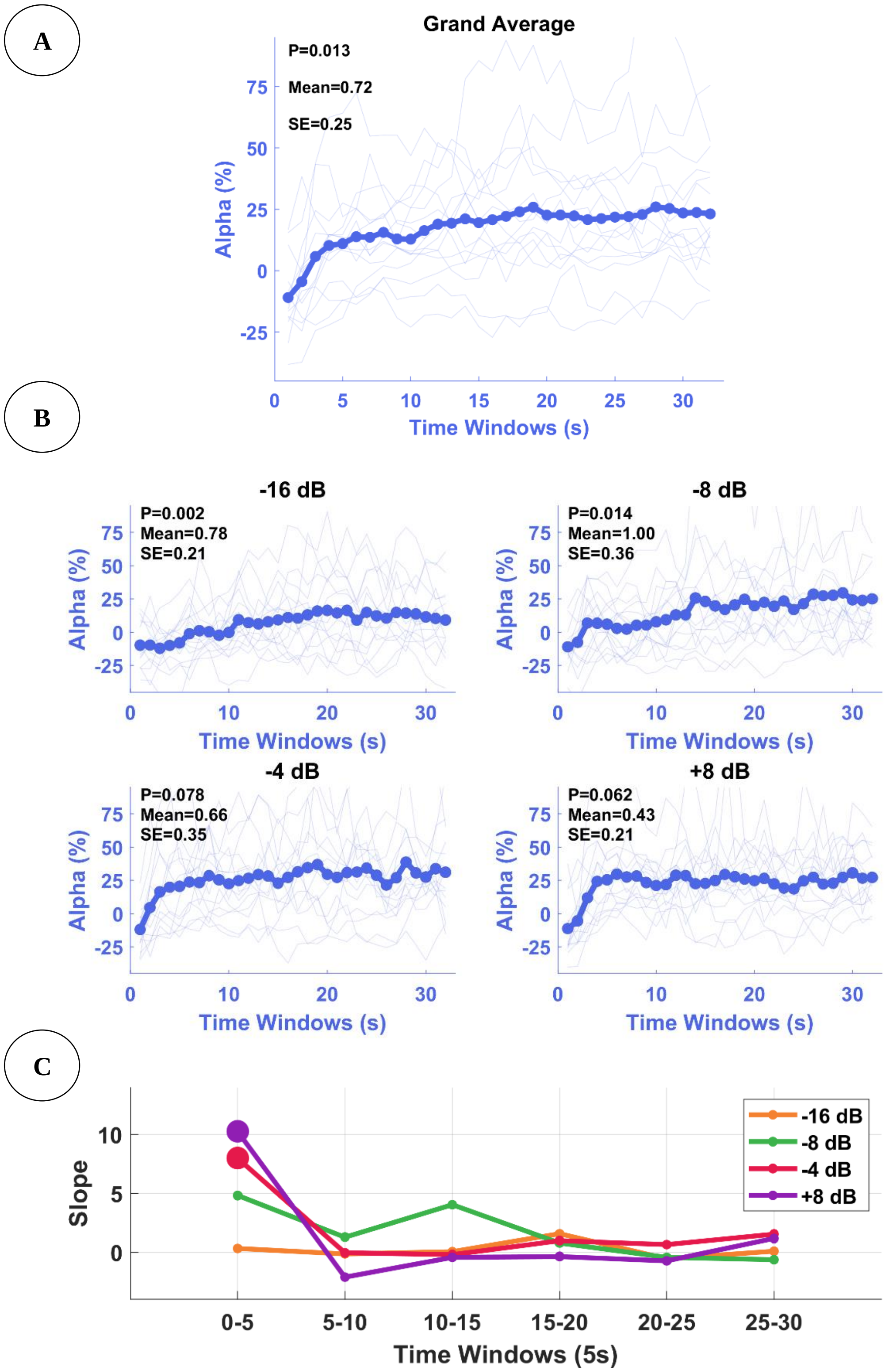


Fig. 5.6 Changes of alpha power using common referencing over time for A) grand average data and B) each condition over 32 seconds. C) The analysis of slope within 5-s intervals showed significantly positive values in the first 5 seconds for +8 and -4 dB conditions (shown in large dots).

Table 5.2 Pearson skipped correlation between performance and EEG using common referencing in the parietal region in different phases. Significant correlations are shown in black.

	Phasic	Tonic		
Band / Electrodes	Alpha	Theta	Alpha	Beta
Parietal	r = 0.03 CI = [-0.20 0.26]	**r = 0.48 CI = [0.31 0.63]**	**r = 0.49 CI = [0.28 0.67]**	**r = 0.46 CI = [0.28 0.61]**

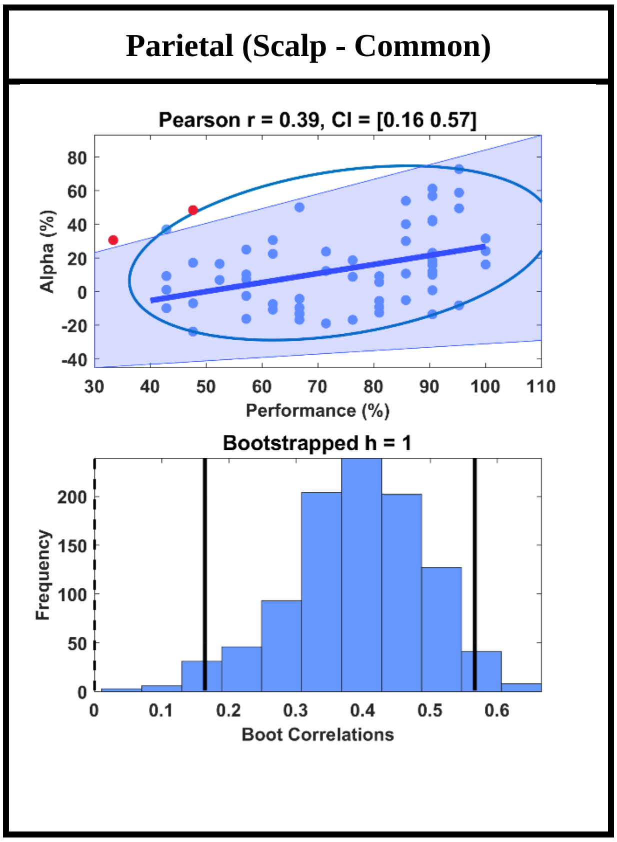


Fig. 5.7. Pearson's skipped correlation between performance and tonic alpha power in the parietal region using common referencing. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

5.3.3 Scalp EEG – Symmetrical

Symmetrical referencing in parietal region (**Table 5.3**) showed similar results to common referencing in the same region. There was no significant change for phasic alpha, but tonic theta, alpha and beta were modulated by SNR and SNR^3. The only difference between common and symmetrical referencing was the modulation of tonic alpha power by SNR^3. The changes in tonic alpha are shown in **Fig. 5.8**.

The analysis of changes over time with symmetrical referencing also showed similar results to common referencing. For the grand average data, there was a significant slope in alpha power over 32 seconds (**Fig. 5.9 top panel**). The two most difficult conditions (-16 and -8 dB) were the only conditions that showed positive slope during 32-second period (**Fig. 5.9 middle panel**). However, unlike common referencing, dividing the alpha power into 5-s windows did not show any significant slope in any of the windows in any condition (**Fig. 5.9 bottom panel**).

The results of skipped Pearson with symmetrical referencing (**Table 5.4**) were also following the same pattern as common referencing. Significant correlations between performance and tonic theta, alpha and beta bands were observed in parietal region. The correlation between tonic alpha and performance is shown in **Fig. 5.10**.

Table 5.3 Results of mixed model based on SNR predictor: estimates of different relative power changes using symmetrical referencing in the parietal region in different bands and phases. Significant p-values are shown in black font.

DF = 57	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = 0.21 SE = 1.07 t = 0.20 p = 0.841	**β = 2.80 SE = 0.86 t = 3.24 p = 0.001**	**β = 2.85 SE = 1.01 t = 2.80 p = 0.006**	**β = 1.89 SE = 0.58 t = 3.24 p = 0.001**
SNR^2	β = 0.02 SE = 0.02 t = 0.95 p = 0.344	β = -0.03 SE = 0.01 t = -1.70 p = 0.092	β = -0.03 SE = 0.02 t = -1.64 p = 0.106	β = -0.007 SE = 0.01 t = -0.57 p = 0.566
SNR^3	β = -0.002 SE = 0.007 t = -0.31 p = 0.756	**β = -0.01 SE = 0.006 t = -2.24 p = 0.028**	**β = -0.01 SE = 0.007 t = -2.03 p = 0.047**	**β = -0.009 SE = 0.004 t = -2.24 p = 0.028**

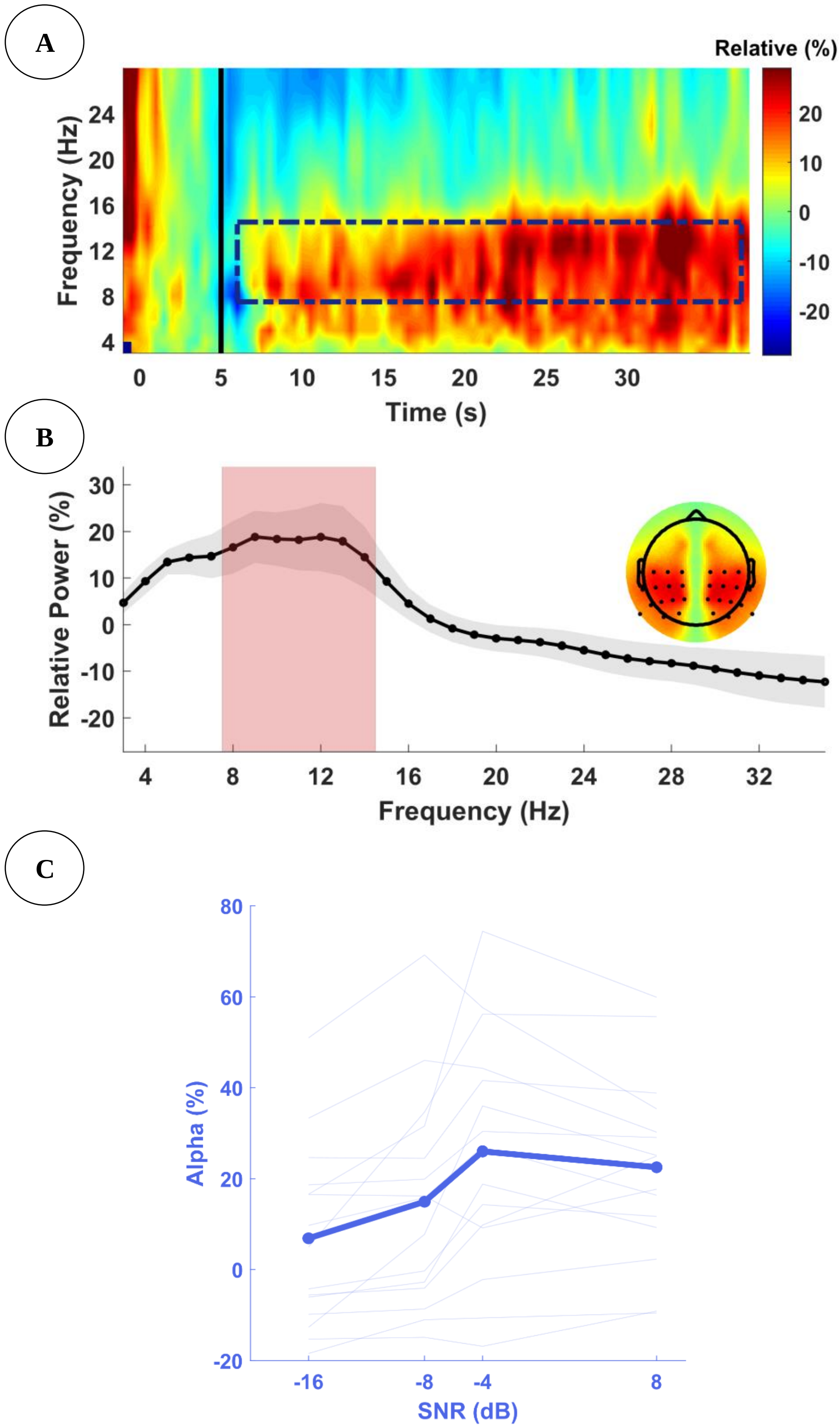


Fig. 5.8 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions after symmetrical referencing. C) The modulation of alpha power by SNR in the highlighted window of panel A which showed significant linear and cubic effects.

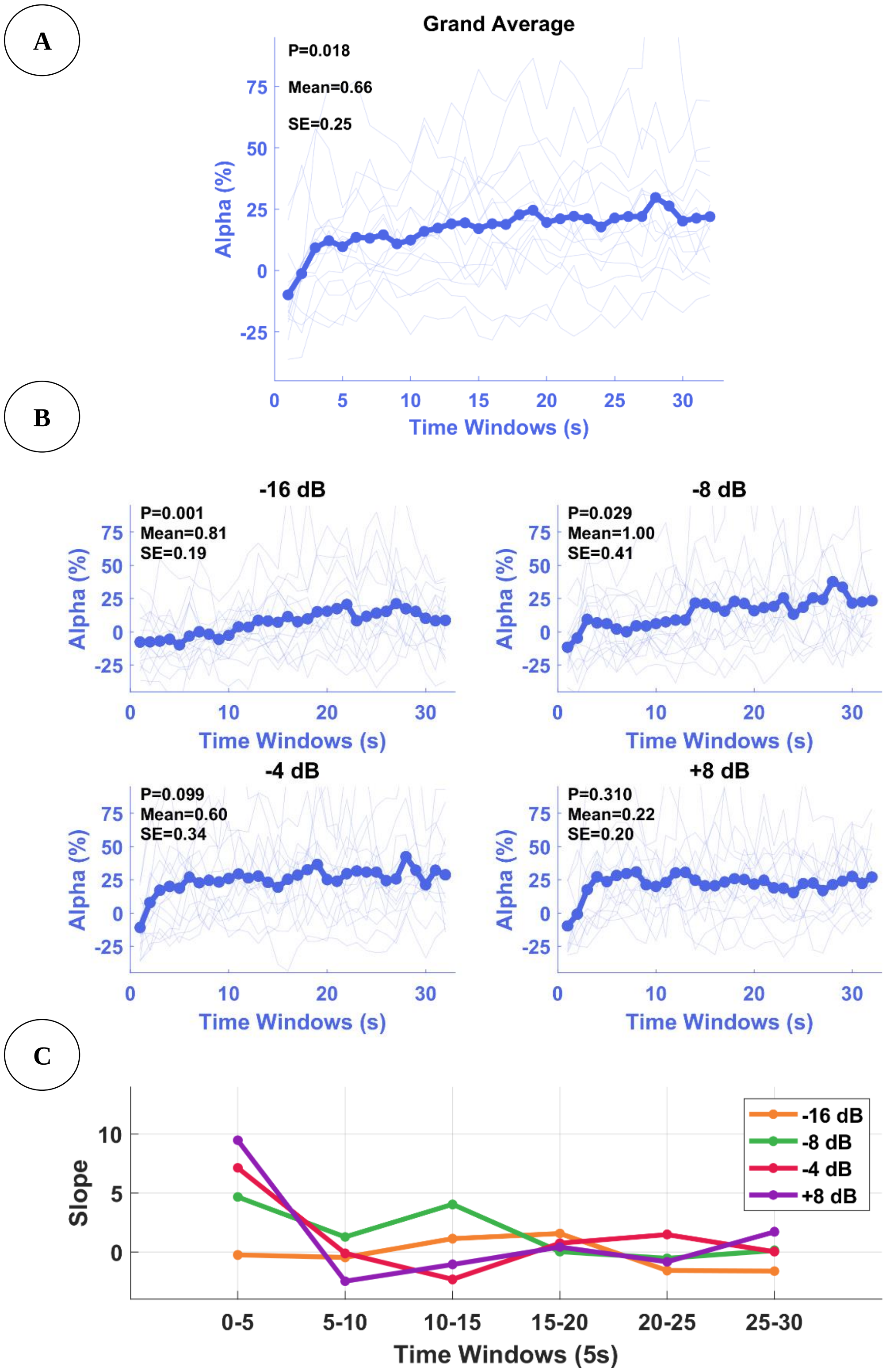


Fig. 5.9 Changes of alpha power using symmetrical referencing over time for A) grand average data and B) each condition over 32 seconds. C) The analysis of slope within 5-s intervals showed no significant positive nor negative values in the first 5 seconds in any condition.

Table 5.4 Pearson skipped correlation between performance and EEG using symmetrical referencing in the parietal region in different phases. Significant correlations are shown in black.

	Phasic	**Tonic**		
Band **Electrodes**	**Alpha**	**Theta**	**Alpha**	**Beta**
Parietal	r = 0.14 CI = [-0.13 0.39]	**r = 0.43** **CI = [0.26 0.60]**	**r = 0.47** **CI = [0.28 0.65]**	**r = 0.29** **CI = [0.06 0.51]**

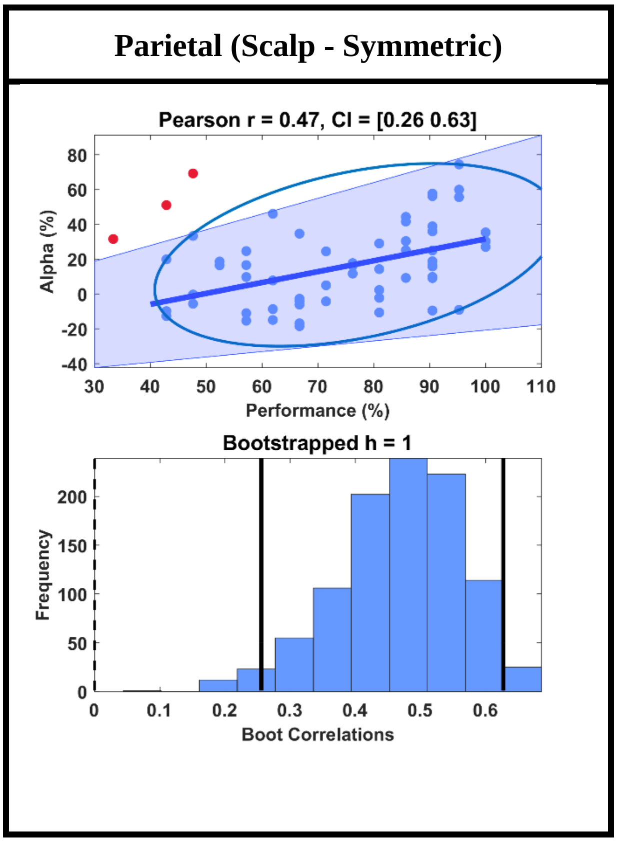


Fig. 5.10. Pearson's skipped correlation between performance and tonic alpha power in the parietal region using symmetrical referencing. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

5.3.4 Ear-EEG

The results of LMM on the ear-EEG data (**Table 5.5**) revealed that only tonic alpha and beta were significantly modulated by SNR and SNR^3. The changes of tonic alpha in ear-EEG are shown in **Fig. 5.11**.

The analysis of changes over time showed no significant slope in the grand average alpha power over 32 seconds (**Fig. 5.12 top panel**). Among all conditions, only the most difficult condition (-16 dB) showed significant positive slope over 32 seconds (**Fig. 5.12 middle panel**). Dividing the alpha power into 5-s windows did not show any significant slope in any of the windows in any condition (**Fig. 5.12 bottom panel**).

Using skipped Pearson on ear-EEG data, tonic alpha power (**Fig. 5.13**) was the only condition that showed significant correlation with performance (**Table 5.6**).

Table 5.5 Results of mixed model based on SNR predictor: estimates of different relative power changes in ear-EEG in different bands and phases. Significant p-values are shown in black font.

DF = 57	Phasic	Tonic		
Band / Predictor	Alpha	Theta	Alpha	Beta
SNR	β = 0.15 SE = 1.10 t = 0.14 p = 0.886	β = 0.74 SE = 0.61 t = 1.21 p = 0.231	**β = 1.90 SE = 0.63 t = 3.00 p = 0.003**	**β = 1.25 SE = 0.42 t = 2.95 p = 0.004**
SNR^2	β = -0.01 SE = 0.02 t = -0.56 p = 0.575	β = -0.002 SE = 0.01 t = -0.22 p = 0.825	β = -0.01 SE = 0.01 t = -1.22 p = 0.226	β = 0.007 SE = 0.009 t = 0.76 p = 0.446
SNR^3	β = -0.001 SE = 0.007 t = -0.15 p = 0.873	β = -0.004 SE = 0.004 t = -0.99 p = 0.325	**β = -0.01 SE = 0.004 t = -2.31 p = 0.024**	**β = -0.007 SE = 0.002 t = -2.54 p = 0.013**

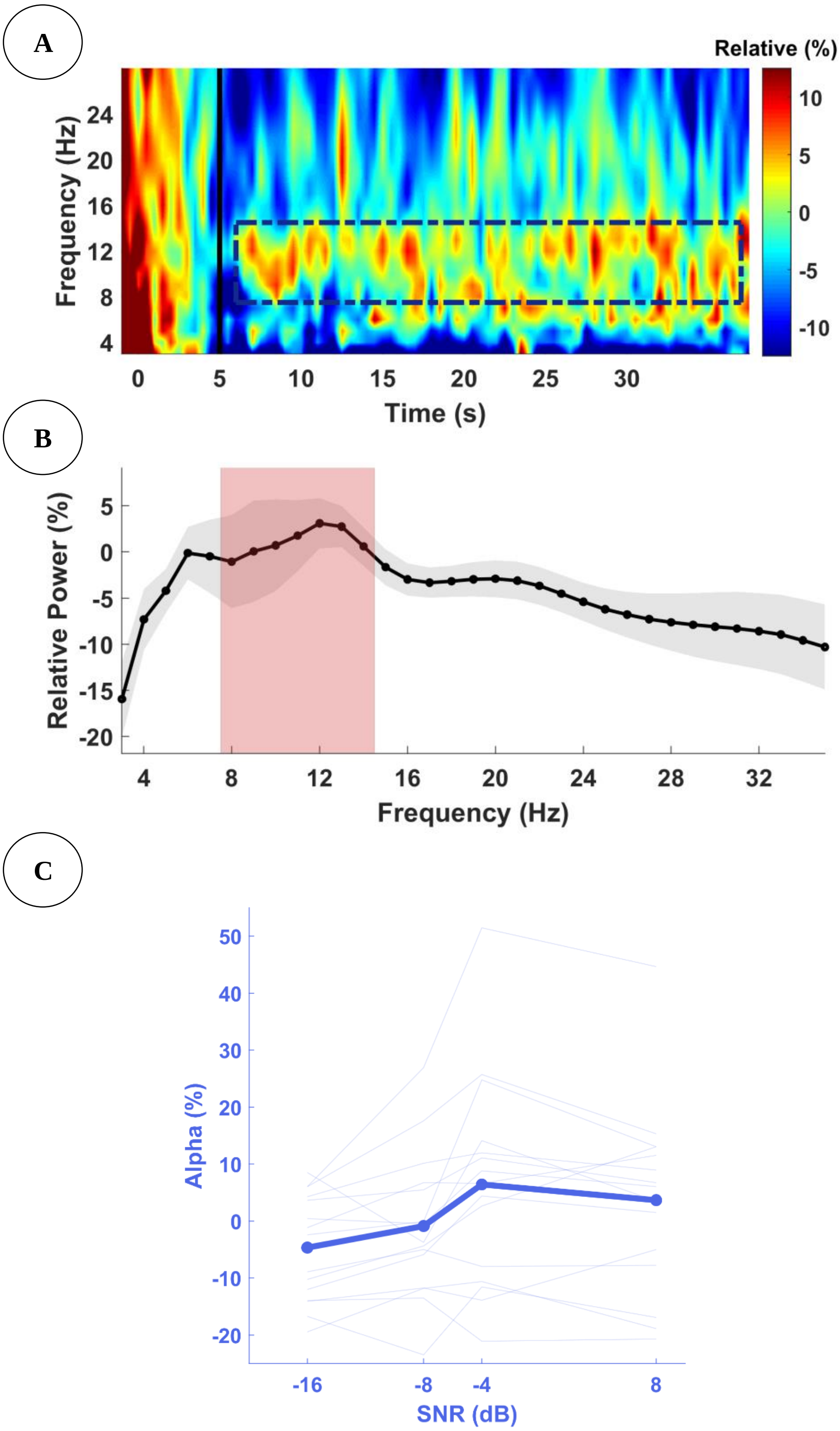


Fig. 5.11 Power changes over time: A) Grand average spectrogram, B) spectrum and topographic map in the highlighted time window of panel A of all participants and conditions in ear-EEG. C) The modulation of alpha power by SNR in the highlighted window of panel A which showed significant linear and cubic effects.

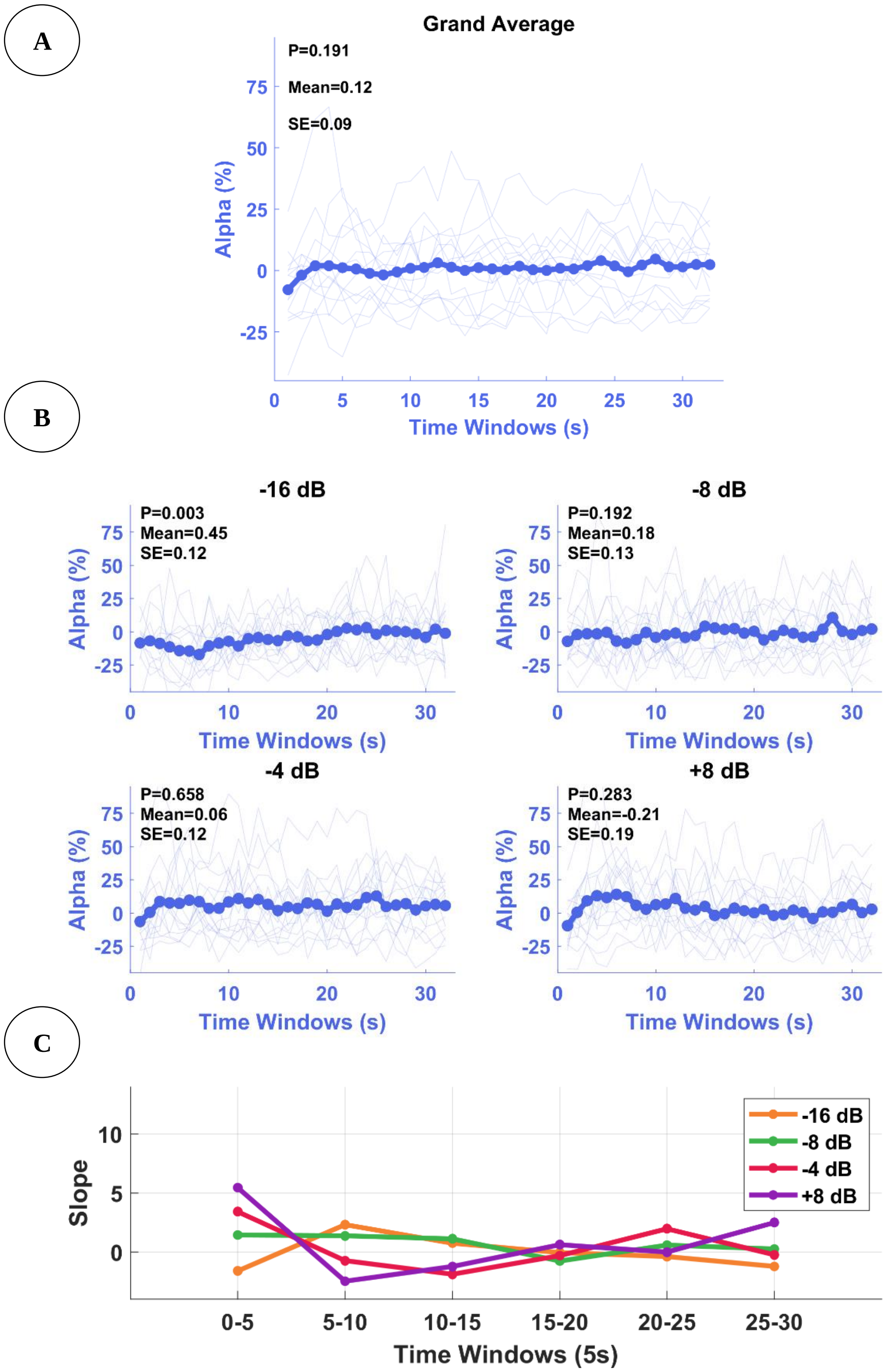


Fig. 5.12 Changes of alpha power (ear-EEG) over time for A) grand average data and B) each condition over 32 seconds. C) The analysis of slope within 5-s intervals showed no significant positive nor negative values in the first 5 seconds in any condition.

Table 5.6 Pearson skipped correlation between performance and ear-EEG in different phases. Significant correlations are shown in black.

	Phasic	Tonic		
Band / Electrodes	Alpha	Theta	Alpha	Beta
Ear-EEG	r = -0.03 CI = [-0.27 0.22]	r = 0.13 CI = [-0.10 0.35]	**r = 0.30** **CI = [0.07 0.51]**	r = 0.03 CI = [-0.23 0.28]

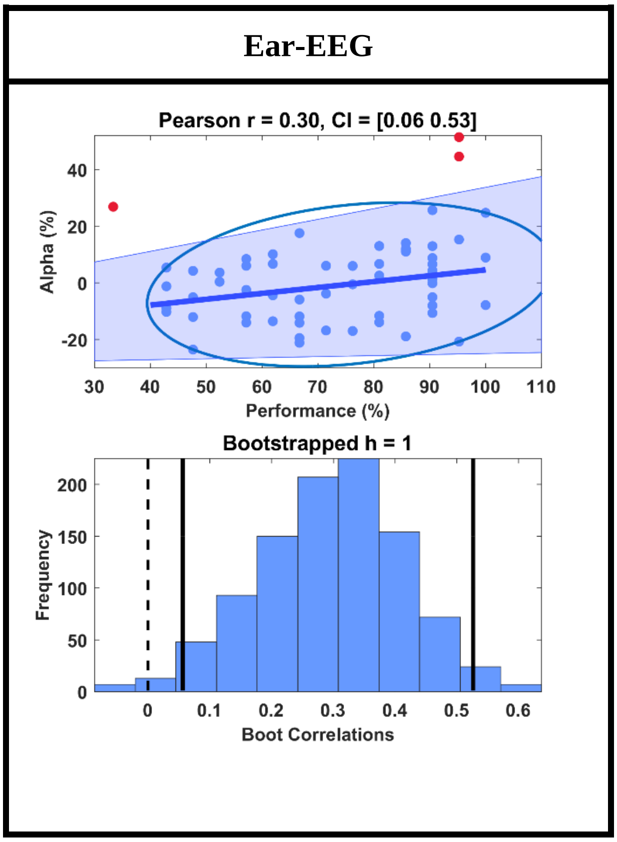


Fig. 5.13. Pearson's skipped correlation between performance and tonic alpha power in ear-EEG. The red dots are considered as outliers by the robust correlation and the shaded area show the 95% CI.

5.4 Discussion

5.4.1 Overview

Using four different SNRs, from very low to very high task demand, alpha power ERS decreased with increasing demand. This was present in both scalp EEG (using common or symmetrical referencing) and ear-EEG. Investigating the changes of alpha power over time in scalp EEG using common referencing showed that less demanding (i.e., more intelligible) conditions had the biggest increase in alpha within the first 5s of the stimuli which led to higher ERS level in those conditions over the full 33s of the stimuli.

Alpha ERS was also positively correlated to performance accuracy in scalp EEG and ear-EEG, which was based on answering questions regarding the contents of the speech.

5.4.2 Changes of EEG power in a continuous speech

One of the advantages of using long, continuous speech is to explore changes of alpha activity over an extended period of listening. Reasonably, it cannot be expected that a listener invests effort continuously and at a constant level during the whole presentation of a continuous speech. A person can adapt to specific listening difficulties, or get fatigued, or lose/gain motivation over time.

Using continuous speech that lasted for 33s, alpha ERD was observed briefly in the beginning of the stimuli (~1 s) which then shifted to alpha ERS. Interestingly, higher SNRs had the biggest leap from alpha ERD to ERS in the first 5s of the stimuli in the highest SNR conditions (-4 and +8 dB). These findings might indicate that successful alpha ERS is necessary for speech understanding in such a continuous, uninterrupted paradigm. The early transition from alpha ERD to ERS might be an early indicator of whether a listening situation is easily intelligible or not.

While similar findings were observed in ear-EEG, they were not statistically significant. This may be an indication that such alpha variations over a continuous speech in different SNRs are more prominent in the central regions of the brain and harder to pick up in or around the ears.

5.4.3 Feasibility of Ear-EEG

Ear-EEG is a feasible device for ambulatory measurement of EEG. While ear-EEG is limited to picking up the electrical brain signals around the ears, it has been shown that it can be successfully used in auditory studies to decode the attended talker in the presence of distracting talkers (Bleichner et al., 2016; Fiedler et al., 2017; Mirkovic et al., 2016). However, to our knowledge no studies have investigated alpha oscillation in continuous speech by using ear-EEG.

The first issue of using ear-EEG in a realistic listening scenario (i.e., continuous speech) is the quality of recorded data compared to the scalp EEG. In this study, for scalp EEG, conductive gel was applied for better contact to the skin, but for ear-EEG the electrodes were used dry. While using conductive gel leads to better-quality signals in general, it is important to implement dry electrodes for any further application of ear-EEGs in real life. Even with the differences in impedance of electrodes, ear-EEG had good quality compared to scalp EEG (on average 1.3 more trials were rejected in ear-EEG compared to scalp EEG), especially in the selected electrodes of the ear-EEG (i.e., K and F).

The second aspect of the ear-EEG results was how the changes of alpha with task demand could be compared in scalp EEG and ear-EEG. It was observed that alpha power changed in a sigmoidal pattern in both scalp EEG (symmetrical referencing) and ear-EEG by manipulating SNR in continuous speech and were correlated to the performance. Similar to scalp EEG, alpha power during such stimuli is probably less a reflection of listening effort and more a measure of performance. More research is required to investigate the feasibility of ear-EEG to measure performance or speech intelligibility, a question that has potential for real-world impact if implemented in hearing aids. To this aim, using real-life scenarios similar to the conversation-like stimuli used in this study is encouraged.

5.4.4 Limitations

The range of implemented SNRs in this experiment varied 24 dB. While this decision of the experiment design was made to investigate a wide range of SNRs, interpreting LMM results on these conditions should be drawn with caution. As the difference between conditions were not equally distributed (8, 4, 12 dB difference between each two consecutive conditions) it might make any non-linear conclusions of the results (such as cubic effect of SNR) spurious.

Another shortcoming was that the scalp and ear-EEG were recorded using separated amplifiers. This decision was made because of hardware restrictions, but it should be considered that the triggers and the ground electrodes were used differently in these measures.

5.5 Conclusion

To our knowledge, this is the first time that oscillations in ear-EEG are investigated to look for changes in listening effort in continuous speech. To do so, we recorded the brain signals simultaneously with dry ear-EEG electrodes with a 64-channel scalp EEG in a continuous speech with four different SNRs to manipulate task demand. The results showed that increased SNR led to increased power activity in theta, alpha and beta bands in both scalp EEG and ear-EEG. It was speculated that such increases were not specific to listening effort and more reflecting performance accuracy.

Chapter 6
General discussion and conclusion

6.1 Overview

According to the FUEL theoretical framework, personal and environmental factors can modify listening effort (Pichora-Fuller et al., 2016). In my thesis, I have investigated this framework in depth, by addressing listening effort using a variety of subjective and objective measurements, and by manipulating the listening task and participants' motivations in a number of ways. Furthermore, given my goal of investigating EEG signals to measure listening effort and understand the neural basis of alpha power, I explored these effects in increasingly ecologically valid situations, and in participants with a variety of hearing abilities.

In order to measure changes in listening effort, I introduced different subjective and objective measures in **Chapter 1**. One of the commonly used objective measures is EEG signals, which pick up electrical activities in the brain. I described one specific feature in EEG signals, alpha power, and how it has been widely used in the literature as a neural correlate of listening effort.

In **Chapter 2**, I examined motivation (a personal factor) and its effects on listening effort, using alpha power. Monetary reward was introduced as a way of manipulating motivation. Task demand (an environmental factor) was also manipulated by varying SNR to create very easy to very difficult listening situations. A traditional speech-in-noise task was used for evaluation of listening, with short sentences as the target audio. I also introduced a maintenance phase where the participants were asked to remember the sentence that they just heard in the presence of background noise. I did not find any effects of monetary reward or SNR on alpha power during the listening phase. However, in the maintenance phase, there was a quadratic interaction between reward and SNR that affected low alpha power: reward led to increased low alpha power when the SNR was not too high and not too low. I drew two conclusions from this experiment. The first conclusion was that during a short-sentence paradigm, low alpha band in the maintenance phase only could be used to evaluate listening effort. When listeners invested more effort, an increase in the alpha power was observed. However, when the task became too difficult, and they started to disengage from the task alpha power decreased. The second conclusion was that reward and SNR interacted with each other to modify effort, as long as task demand was not extremely high or low.

In **Chapter 3**, I used a similar design, but this time focused only on environmental factors, testing whether reverberation, using different room simulations, can affect listening effort. Similar to the results of **Chapter 2**, low alpha power in the maintenance phase was modulated by the interaction of SNR and room simulations. In this study, I also used questionnaires as a subjective measure of listening effort to investigate whether subjective and objective measures of listening effort are co-modulated or correlated to each other. While the low alpha power in the maintenance phase (i.e., an objective measure of listening effort) was modified in an inverted U-shaped form with task demand, self-reported effort (i.e., a subjective measure of listening effort) showed a linear increase in ratings with increasing task demand. I drew three conclusions from this experiment. The first conclusion was that, once again, the power of low alpha band could be utilized in the maintenance phase as a measure of listening effort in a short-sentence paradigm: when listeners invested more effort, increase in alpha power was observed until the task became too difficult and they started to disengage from the task which led to decrease in alpha power (more discussion in **Section 6.2**). The second conclusion was that different rooms (with different reverberation) could influence listening effort depending on the SNR. The third conclusion was that subjective and objective measures of listening effort did not show similar patterns. While the subjective measure changed linearly with task demand, the objective measure formed an inverted U-shaped curve with changes of task demand. I speculate the inconsistency between subjective and objective measures only happens when listening effort varies in a wide range of task difficulties which can cover a wide range of psychometric function of speech intelligibility. It is possible that for speech intelligibilities higher than the middle point of the psychometric function (50% speech intelligibility), both subjective and objective measures of effort agree with each other. This is because the more demanding the task gets the more effort listeners invest in the task and this is reflected by both subjective perception and physiological changes. However, for the speech intelligibilities below the middle point of psychometric function, subjective measures indicate that effort increases, while objective measures indicate that effort decreases. The physiological measures suggest that listeners start to disengage from the task even while listeners feel that they are putting in additional effort. This shows the perception of putting more effort is not similar to physiological manifestation of effort. It is possible that not enough sensory information is delivered to the brain for processing of the information and thus less manifestation of effort is observed by exploring physiological changes, but the listeners still perceive they put in additional effort.

In **Chapter 4**, I decided to move away from normal-hearing participants and the artificial short-sentence paradigm of chapters 1-3, which is quite different to everyday listening situations. I instead focused on longer, realistic monologue stimuli and how people with hearing impairment apply effort while listening to long stimuli. I used these longer stimuli from actual newscasts to see if prior effects held in a context that would be more generalisable to everyday life. Therefore, in **Chapter 4**, I introduced a continuous-speech paradigm lasting over 30 s and conducted two experiments on hearing impaired participants. In the first experiment hearing impaired participants (using hearing aids, but without any manipulations) were exposed to two different SNRs. The alpha changes were opposite to what we observed in the previous chapters: when listeners invested more effort during the listening, alpha power decreased. In the second experiment, SNR as well as the noise reduction scheme in hearing aids (on vs. off) were manipulated. While I did not observe any effects of noise reduction on alpha power, the condition requiring more effort (lower SNR) led to decreased alpha power. These findings were opposite to what I found in **Chapter 2** and **Chapter 3** where more effort led to more alpha power, potentially due to different tasks which might require different processing in the brain. While the task in short-sentence paradigms were to repeat the words, the task in continuous speech was to get the gist of the speech. It is possible that alpha power during continuous speech is more comparable to the listening phase in shorter speech paradigm and both are as results of stimulus-driven responses. This might be one explanation that why there was a significant correlation between alpha power in the continuous speech and performance.

In **Chapter 5**, I intended to test the effects of continuous speech on normal-hearing participants to replicate the effects of lower alpha power being greater effort, as the results in **Chapter 4** could have been due to the different population. I also explored whether the changes of alpha power in this paradigm can be picked up by an ambulatory measurement such as ear-EEG. To this end, I had a similar design to the studies in **Chapter 4**, but this time the experiment was conducted on normal hearing individuals and ear-EEG was recorded simultaneously with scalp EEG. The participants listened to a continuous speech with four varying SNRs while scalp EEG and ear-EEG were simultaneously recorded. In line with the results of **Chapter 4**, I observed that with more effort, alpha power was decreased. This modulation could be detected both by scalp and ear-EEG in the alpha band. I drew two conclusions from this study. The first conclusion was that ear-EEG is capable of detecting alpha changes during listening to a continuous-speech paradigm. The second, and more general conclusion based on the results of **Chapter 4** and **Chapter 5**, was that in a continuous-speech paradigm alpha power might be reflecting something other than listening effort. Since alpha power and performance directly changed with changes of SNR (i.e., higher SNR led to higher alpha power and performance), it is possible that alpha power in continuous speech reflected performance (more discussion in **Section 6.3**).

6.2 Alpha during short sentences

Based on the results of the two experiments in **Chapter 2** and **Chapter 3**, I did not find evidence that alpha power reflected changes of effort during the listening phase, while listening to a short sentence (~1 s). The alpha ERD during the listening phase was not modulated by SNR or reward (**Chapter 2**) or by room simulations (**Chapter 3**), but it changed monotonically to SNR manipulations (**Chapter 3**). These results do not support some previous works that have shown alpha changes during listening to short sentences can be used as a measure of listening effort (e.g., Obleser et al., 2012; Wöstmann et al., 2015) or can be used to predict speech intelligibility (e.g., Obleser & Weisz, 2012). However, during the maintenance phase, low alpha power ERS was modulated by an interaction of SNR and reward (**Chapter 2**) and SNR and room simulations (**Chapter 3**) in an inverted U-shaped form. This complements the observed changes in alpha power related to task demand reported by numerous auditory and non-auditory studies during maintenance, (e.g., Jensen et al., 2002; Obleser et al., 2012; Tuladhar et al., 2007; Wisniewski et al., 2015, 2017).

In **Chapter 1**, I discussed how alpha activation is considered as "functional inhibition" (Jensen & Mazaheri, 2010; Klimesch, 2012; Klimesch et al., 2007). During an effortful task, the role of alpha oscillation could be to inhibit the task-irrelevant regions, so the information is gated to the relevant areas of the brain. Given that during the maintenance phase of our studies background noise was present, it seems logical that alpha activation served as "functional inhibition". However, since the background noise was always constant, the changes of alpha power could not have related only to processing of that background noise. Instead, the changes may have related to continued processing of the speech during the maintenance phase, with the difficulty of processing that speech leading to different levels of alpha power being required to inhibit background noise during maintenance. This could have been due to the fact that more demanding listening situations led to more noisy representations of words and more mismatches between the representation of words in episodic memory and semantic memory. The greater the strain on working memory during listening, the greater the effort required during the maintenance phase, evident through alpha oscillation. However, it is unclear why alpha power did not change during listening with the changes of SNR and reward (**Chapter 2**) and reverberation time (**Chapter 3**). One possible explanation is that there might have been other alpha oscillations in other regions of the brain with different roles (for example possibly due to spatial attention) that cancelled out "effortful" alpha oscillation in EEG signals. It is important to note that alpha power showed a general desynchronization (i.e., ERD) in the short period of listening phase which was only significant with the manipulation of SNR (**Chapter 3**), but it did not reflect any inverted U-shaped pattern. However, alpha power showed general synchronization (i.e., ERS) in the maintenance phase which changed in an inverted U-shaped pattern with task demand (**Chapter 2** and **Chapter 3**) and motivation (**Chapter 2**). This might suggest that alpha ERD is more of a low-level response to auditory stimuli, whereas alpha ERS is probably more high level and can change based on the difficulty of the task.

6.3 Alpha during continuous speech

Based on the results of the three experiments in **Chapter 4** and **Chapter 5,** in which I used continuous speech stimuli, I concluded that alpha power is modulated by SNR, but it is plausible that this does not reflect listening effort. I drew this conclusion based on two observations.

The first observation was that alpha power decreased with increasing task demand. This was not in line with the "functional inhibition" theories on alpha power. Based on these theories if task demand is higher, then alpha power should be increased in order to inhibit irrelevant regions of the brain to gate the information to more relevant regions to improve the SNR of the sound of interest (Jensen & Mazaheri, 2010; Klimesch, 2012; Klimesch et al., 2007). However, I observed that during continuous speech, lower SNR (i.e., more demanding) in fact led to decreased alpha power.

The second observation was that the changes of alpha power were correlated to the performance of the participants. The performance in those continuous-speech experiments was measured for each trial by answering a single question regarding the contents of the speech. While this is a crude measurement of speech intelligibility, it was still affected significantly by the level of SNR: higher SNR led to more accuracy in responses. That is, an increase in SNR led to an increase in both performance and alpha power. While it is impossible to draw any conclusions on the causality between these measures (i.e., did higher SNR lead to more alpha power and thus better performance or did higher SNR lead to better performance and thus higher alpha power?), one plausible conclusion is that alpha power changed with performance. In an unexpected finding, different talkers (one spoken by a male and one by a female) led to significant changes in performance and alpha power in the second study of **Chapter 4.** While this was not controlled for, it still showed that alpha power is not merely an SNR tracker in a continuous speech and might as well reflect performance.

If alpha power in continuous speech can reflect performance, then it may indirectly be used to measure speech intelligibility as well. The problem is that in continuous designs such as ours, it is difficult to have a perfect measure of speech intelligibility (one may hear all the words spoken over 30 s, but hardly remember all the words to repeat for a valid speech intelligibility score). It remains to be seen if future studies can overcome this obstacle and investigate the relation between alpha power and speech intelligibility in such a continuous-speech paradigm.

6.3.1 Feasibility of ear-EEG

Based on the results of the experiment in **Chapter 5**, alpha changes could be measured in ear-EEG by varying SNR in continuous speech. While ear-EEG has many limitations compared to scalp EEG (including fewer number of electrodes and limited spatial resolution), these results are promising for future advances in hearing aids. The electrodes used for ear-EEG were dry electrodes that needed no extra preparation including applying conductive gel for better contact with the skin. That makes dry ear-EEG suitable for implementation in hearing aids.

Not much work has been done on evaluation of listening effort with ear-EEG. While previous studies have shown that ear-EEG can be used in auditory tasks for different purposes such as attention decoding (e.g., Fiedler et al., 2017; Mirkovic et al., 2016), more studies are required to look for applications of ear-EEG. Especially during listening to continuous speech which can often happen in daily life, ear-EEG can be implemented in hearing aids which can provide valuable information regarding alpha power. Based on the results of **Chapter 4** and **Chapter 5**, alpha power in ear-EEG during continuous speech was increased with increasing SNR which can be used as a measure of performance during listening. However, the results of the study are based on group-level analyses and more individualized approaches are needed for real-life applications. Identifying individual alpha frequency (IAF) for each user (e.g., Klimesch, 1999) may help towards using ear-EEG in individual level. A possible application for ear-EEG is to be implemented within a hearing aid to help steer the signal processing towards the hearing-aid user's intent. By using machine learning techniques and training on ear-EEG data, useful features can be extracted from the EEG signals of the user in order to automatically adjust a hearing aid setting in different listening situations for better hearing and/or reduced listening effort. While in this thesis, power in the alpha band may have been a measure for performance during continuous speech, there have been an increasing number of studies that showed ear-EEG in other bands can be used to decode attention when there are several sources of sound that are spatially separated from each other (e.g., Alickovic et al., 2020; Bleichner et al., 2016). My work indicates the feasibility of using ear-EEG for extracting alpha power in continuous speech to measure performance depending on the task at hand. While promising, the potential of ear-EEG, especially within hearing aids in real-life situations, requires more research in future.

6.4 Ecological validity

I investigated the changes of alpha power during effortful listening in both short and continuous speech. As I found different patterns of alpha power, and possibly different roles of alpha power, between short and continuous speech, it is vital to revisit the main question of this thesis more closely. The original aim of this thesis was to investigate whether EEG signals (mainly alpha power) reflect listening effort in ecologically valid situations.

I studied a variety of conditions to draw conclusions about ecologically valid listening: in **Chapter 2**, I investigated the impact of different levels of motivation, and in **Chapter 3,** I investigated the impact of different room types. However, while short-sentence paradigms such as those conducted in **Chapter 2** and **Chapter 3** allowed us to control for speech intelligibility and study EEG signals across a wide range of task demands, these short, single sentences barely happen in real life, especially with those added “artificial” maintenance phases. Using this standard paradigm helped me to investigate alpha changes related to different manipulations while measuring speech intelligibility (e.g., SNR, reward, and reverberation), but the cost of this level of experimental control was that I had to remove other variables from the experimental environment, potentially reducing generalisability to real-life scenarios (Keidser et al., 2020). Therefore, I decided to move towards more realistic situations using continuous speech in **Chapter 4** and **Chapter 5** which led to some fundamental differences in the study designs. For example, speech intelligibility could not be controlled or measured anymore. Maybe more importantly, the nature of the task was different. While in short-sentence paradigms participants have to memorise every single word without any context, in continuous-speech paradigms they are not required to memorise any words but instead get the gist of the speech. Another important difference between the two paradigms is that in short-speech paradigms the listening and maintenance phases are separated from each other, however in the continuous speech paradigm, listening and maintenance phases are entangled. These differences all may have played a part in the conflicting changes of alpha power in continuous speech compared to short sentences.

Given the more natural listening task used in the later chapters, it might be most sensible to focus on **Chapter 4** and **Chapter 5** to address the main question of this thesis: How can alpha power be used to objectively measure listening effort in ecologically valid situations? My conclusion is that in such real-life situations, alpha power may not be a direct measure of listening effort but may instead reflect performance. However, it should be noted that ecological validity is not only limited to presenting continuous speech and many more parameters are involved in listening in real life. All of these show the complexity of studies on listening effort in ecological situations and that more research is required for us to answer how we can measure listening effort more consistently and reliably.

6.5 Limitations and future direction

As mentioned in **Section 6.4**, the changes observed in alpha power due to manipulations of reward in **Chapter 2** and reverberation time in **Chapter 3** were during an artificially added maintenance phase where the target has ended but the noise continues. This allowed the measurement not to be confounded by the offset of the target. Also, the addition of maintenance phase was necessary to measure alpha power during a fixed SPL in order to show that the changes of alpha power was due to effortful listening to the target speech and not the overall SPL. However, during our daily lives, we are often exposed to either continuous speech or short sentences in a conversation that come constantly after each other without much time for *pure maintenance* phase. Therefore, such short-sentence paradigm with maintenance phase rarely happens in real life. More research is required to look for the effects of motivation and reverberation during more realistic listening scenarios, such as using continuous speech with those manipulations.

One of the limitations of the studies with continuous speech in **Chapter 4** and **Chapter 5** was our inability to measure speech intelligibility during the task. This is important because based on the results of **Chapter 3**, depending on the speech intelligibility level, alpha power may increase or decrease with increasing task demand. Therefore, speech intelligibility can help us to understand whether the listeners are trying to spend more effort or starting to disengage from the task. One way to overcome this issue in future is to ask the participants to estimate their own speech intelligibility. As shown in **Chapter 3**, test participants could have an accurate estimation of their own speech intelligibility. While it might be more difficult to estimate one’s subjective intelligibility during a longer stimulus, the information can still be useful in further interpretation of alpha power.

Another limitation was that the only acoustic change during the continuous speech was SNR. The reason that I speculated that alpha power is not merely a SNR tracker is because of the results of **Chapter 4**, where an unanticipated difference between the two talkers (one male and one female) proved to change the alpha power significantly. As this was a strong effect, it suggests that features of the voice are more important than I anticipated. Given voices differ on features like saliency, a next step could be to control for other acoustic features such as saliency to test whether alpha power can reflect those changes as well. Therefore, it is important to have more controlled acoustic parameters during continuous speech for better evaluation of alpha power.

Last but not least, EEG data contains large amounts of information on different ongoing processing in the brain, as well as physiological and environmental artefacts (such as eye movement or heart rate variability which may be useful at times for task evaluation). The main focus of this thesis was on the changes of alpha power, which is just one of many features that can be extracted from the EEG signals (albeit a very important one). The challenge of studying alpha power is that it can change with fatigue or drowsiness during a cognitive task. As seen in **Chapter 4** and **Chapter 5**, alpha power is also dynamic and can change over time. This can be an advantage towards understanding the cognitive aspects of fatigue, as well as potentially serve as a measure of listening engagement. Nevertheless, the changes in alpha power with changes in the realism of the task in these chapters urge caution on what conclusions should be drawn with any given conditions tested.

6.6 Conclusion

The aim of this thesis was to investigate how alpha power of EEG signals can be used as an objective measure of listening effort in ecologically valid situations. Two different listening paradigms were used: short and continuous speech. After listening to short sentences, I found evidence suggesting that low alpha power in the maintenance phase reflects listening effort. I observed that personal factors (such as motivation which was manipulated by reward) and environmental factors (such as different room simulations characterised through reverberation time) modified listening effort as a result of interacting with task demand (by varying SNR): when participants invested more effort, alpha power was increased until task demand became too high and participants started to disengage from the task and thus alpha power started to decrease (i.e., inverted U-shaped curve). However, using continuous speech, alpha power during listening decreased with increasing task demand (decreasing SNR). Different patterns of alpha power in short and continuous speech indicate various roles of alpha power when the speech is a single, short sentence versus when it is long, and uninterrupted. While it is possible that alpha power in the maintenance phase of short speech reflects listening effort, it would appear that in during listening to a continuous speech, alpha power is not a direct measure of listening effort and instead reflects performance of listening.

List of References

Adrian, & Matthews. (1934). The Berger rhythm: potential changes from the occipital lobes in man. *Brain*, *57*(4), 355–385. http://brain.oxfordjournals.org/

Ahn, J. W., Ku, Y., Kim, D. Y., Sohn, J., Kim, J. H., & Kim, H. C. (2018). Wearable in-the-ear EEG system for SSVEP-based brain–computer interface. *Electronics Letters*, *54*(7), 413–414. https://doi.org/10.1049/el.2017.3970

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000697

Alickovic, E., Lunner, T., Wendt, D., Fiedler, L., Hietkamp, R., Ng, E. H. N., & Graversen, C. (2020). Neural Representation Enhanced for Speech and Reduced for Background Noise With a Hearing Aid Noise Reduction Scheme During a Selective Attention Task. *Frontiers in Neuroscience*, *14*, 846. https://doi.org/10.3389/fnins.2020.00846

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—a review. *International Journal of Audiology*, *42 Suppl 2*, 2S17-20. https://doi.org/10.3109/14992020309074639

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Barone, J., & Rossiter, H. E. (2021). Understanding the Role of Sensorimotor Beta Oscillations. *Frontiers in Systems Neuroscience*, *0*, 51. https://doi.org/10.3389/FNSYS.2021.655886

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berger, H. (1929). Über das elektroenkephalogramm des menschen. *Arch Psychiatr Nervenkr*, *87*(1), 527–570.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., & Corona-Strauss, F. I. (2017). Neurodynamic evaluation of hearing aid features using EEG correlates of listening effort. *Cognitive Neurodynamics*, *11*(3), 203–215. https://doi.org/10.1007/s11571-017-9425-5

Bleichner, M. G., Lundbeck, M., Selisky, M., Minow, F., Jäger, M., Emkes, R., Debener, S., & De Vos, M. (2015). Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? *Physiological Reports*, *3*(4), e12362. https://doi.org/10.14814/phy2.12362

Bleichner, M. G., Mirkovic, B., & Debener, S. (2016). Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *Journal of Neural Engineering*, *13*(6). https://doi.org/10.1088/1741-2560/13/6/066004

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Current Biology*, *22*(20), 1969–1974. https://doi.org/10.1016/j.cub.2012.08.029

Brehm, J. W., & Self, E. A. (1989). The Intensity of Motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Cabestrero, R., Crespo, A., & Quirós, P. (2009). Pupillary dilation as an index of task demands. *Perceptual and Motor Skills*, *109*(3), 664–678. https://doi.org/10.2466/PMS.109.3.664-678

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press Journals.

Cvijanović, N., Kechichian, P., Janse, K., & Kohlrausch, A. (2017). Effects of noise on arousal in a speech communication setting. *Speech Communication*, *88*, 127–136. https://doi.org/10.1016/j.specom.2017.02.001

Darwin, C. (1872). The expression of emotion in animals and man. *London, England: Murray*.

Das, N., Bertrand, A., & Francart, T. (2018). EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *Journal of Neural Engineering*, *15*(6), 066017. https://doi.org/10.1088/1741-2552/aae0a6

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932. https://doi.org/10.1162/jocn_a_00084

de Cheveigné, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage*, *172*, 903–912. https://doi.org/10.1016/j.neuroimage.2018.01.035

De Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. In *NeuroImage* (pp. 98:487-505). https://doi.org/10.1016/j.neuroimage.2014.05.068

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592096511. https://doi.org/10.1177/2515245920965119

Decruy, L., Lesenfants, D., Vanthornhout, J., & Francart, T. (2020). Top-down modulation of neural envelope tracking: the interplay with behavioral, self-report and neural measures of listening effort. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.14753

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desarnaulds, V., Carvalho, A. P. O., & Monay, G. (2002). *Church Acoustics and the*

Influence of Occupancy. *9*(1), 29–47.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical Alpha Oscillations Predict Speech Intelligibility. *Frontiers in Human Neuroscience, 11*, 88. https://doi.org/10.3389/fnhum.2017.00088

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*. https://doi.org/10.1038/s41598-019-47643-1

Dockree, P. M., Kelly, S. P., Foxe, J. J., Reilly, R. B., & Robertson, I. H. (2007). Optimal sustained attention is linked to the spectral content of background EEG activity: Greater ongoing tonic alpha (∼10 Hz) power supports successful phasic goal activation. *European Journal of Neuroscience*, *25*(3), 900–907. https://doi.org/10.1111/j.1460-9568.2007.05324.x

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is listening in noise worth it? the neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing*, *37*(Suppl 1), 101S-110S. https://doi.org/10.1097/AUD.0000000000000300

Enriquez-Geppert, S., Huster, R. J., & Herrmann, C. S. (2017). EEG-neurofeedback as a tool to modulate cognition and behavior: A review tutorial. *Frontiers in Human Neuroscience*, *11*, 51. https://doi.org/10.3389/fnhum.2017.00051

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, *14*(3). https://doi.org/10.1088/1741-2552/aa66dd

Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology*, *0*(JUL), 154. https://doi.org/10.3389/FPSYG.2011.00154

Gatehouse, S., & Akeroyd, M. A. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology*, *45 Suppl 1*(SUPPL. 1). https://doi.org/10.1080/14992020600783103

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*. https://doi.org/10.1093/cercor/7.4.374

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. In *Neurosurgery Clinics of North America* (Vol. 22, Issue 2, pp. 133–139). NIH Public Access. https://doi.org/10.1016/j.nec.2010.11.001

Granholm, E., Asarnow, R., Sarkin, A., & Dykes, K. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. https://doi.org/10.1111/J.1469-8986.1996.TB01071.X

Hagerman, B., & Olofsson, Å. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica United with Acustica*, *90*(2), 356–361.

Hallberg, L. R.-M., & Carlsson, S. G. (1993). A qualitative study of situations turning a hearing disability into a handicap. *Disability, Handicap & Society, 8*(1), 71–86. https://doi.org/10.1080/02674649366780051

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. https://doi.org/10.7717/peerj.4794

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hartmann, T., Lorenz, I., Müller, N., Langguth, B., & Weisz, N. (2014). The effects of neurofeedback on oscillatory processes related to tinnitus. *Brain Topography*, *27*, 149–157. https://doi.org/10.1007/s10548-013-0295-9

Hasselmo, M. E., & Stern, C. E. (2014). Theta rhythm and the encoding and retrieval of space and time. *NeuroImage*, *85*(0 2), 656. https://doi.org/10.1016/J.NEUROIMAGE.2013.06.022

Hauswald, A., Keitel, A., Chen, Y., Rösch, S., & Weisz, N. (2020). Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *European Journal of Neuroscience*, *00*, 1–15. https://doi.org/10.1111/ejn.14912

Hazrati, O., & Loizou, P. C. (2012). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International Journal of Audiology*, *51*(6), 437–443. https://doi.org/10.3109/14992027.2012.658972

Hidi, S. (2016). Revisiting the Role of Rewards in Motivation and Learning: Implications of Neuroscientific Research. *Educational Psychology Review*, *28*, 61–93. https://doi.org/10.1007/s10648-015-9307-5

Hjortkjaer, J., Märcher-Rørsted, J., Fuglsang, S. A., & Dau, T. (2018). Cortical oscillations and entrainment in speech processing during working memory load. *European Journal of Neuroscience*, *June 2017*, 1–11. https://doi.org/10.1111/ejn.13855

Hjortkjær, J., Märcher-Rørsted, J., Fuglsang, S. A., & Dau, T. (2020). Cortical oscillations and entrainment in speech processing during working memory load. *European Journal of Neuroscience*, *51*(5), 1279–1289. https://doi.org/10.1111/ejn.13855

Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. I. (2009). Estimating brain load from the EEG. *TheScientificWorldJournal*, *9*, 639–651. https://doi.org/10.1100/tsw.2009.83

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

Houben, R., Van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of*

Audiology, *52*(11), 753–761. https://doi.org/10.3109/14992027.2013.832415

Hughes, S. E., Rapport, F., Watkins, A., Boisvert, I., McMahon, C. M., & Hutchings, H. A. (2019). Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI). *BMJ Open*, *9*(7), e028881. https://doi.org/10.1136/BMJOPEN-2018-028881

Jaworski, A., & Stephens, D. (1998). Self-reports on silence as a face-saving strategy by people with hearing impairment. *International Journal of Applied Linguistics (United Kingdom)*, *8*(1), 61–80. https://doi.org/10.1111/j.1473-4192.1998.tb00121.x

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex (New York, N.Y. : 1991)*, *12*(8), 877–882.

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*, 186. https://doi.org/10.3389/fnhum.2010.00186

Johnson, J., Xu, J., Cox, R., & Pendergraf, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American Journal of Audiology*, *24*(3), 419–431. https://doi.org/10.1044/2015_AJA-14-0058

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: PrenticeHall, Inc.

Kaiser, J., & Lutzenberger, W. (2003). Induced gamma-band activity and human brain function. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *9*(6), 475–484. https://doi.org/10.1177/1073858403259137

Kappel, S. L., Rank, M. L., Toft, H. O., Andersen, M., & Kidmose, P. (2019). Dry-Contact Electrode Ear-EEG. *IEEE Transactions on Biomedical Engineering*, *66*(1), 150–158. https://doi.org/10.1109/TBME.2018.2835778

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. In *Neuroscientist* (Vol. 20, Issue 5, pp. 509–521). SAGE Publications Inc. https://doi.org/10.1177/1073858413514136

Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 Prescription Procedure. *Audiology Research*, *1*(1), 88–90. https://doi.org/10.4081/AUDIORES.2011.E24

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., & Smeds, K. (2020). The Quest for Ecological Validity in Hearing Science: What It Is, Why It Matters, and How to Advance It. *Ear and Hearing*, *41*, 5S-19S. https://doi.org/10.1097/AUD.0000000000000944

Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., & Mandic, D. P. (2013). A study of evoked potentials from ear-EEG. *IEEE Transactions on Biomedical Engineering*, *60*(10), 2824–2830. https://doi.org/10.1109/TBME.2013.2264956

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. In *Brain Research Reviews* (pp. 53(1):63-88). https://doi.org/10.1016/j.brainresrev.2006.06.003

Klimesch, W., Schimke, H., & Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. *Brain Topography*, *5*(3), 241–251. https://doi.org/10.1007/BF01128991

Knecht, H. A., Nelson, P. B., Whitelaw, G. M., & Feth, L. L. (2002). Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *American Journal of Audiology*, *11*(2), 65–71. https://doi.org/10.1044/1059-0889(2002/009)

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291–300. https://doi.org/10.1097/AUD.0B013E3182310019

Koelewijn, T., Zekveld, A. A., Lunner, T., & Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106–112. https://doi.org/10.1016/j.heares.2018.07.011

Kostandyan, M., Bombeke, K., Carsten, T., Krebs, R. M., Notebaert, W., & Boehler, C. N. (2019). Differential effects of sustained and transient effort triggered by reward – A combined EEG and pupillometry study. *Neuropsychologia*, *123*, 116–130. https://doi.org/10.1016/j.neuropsychologia.2018.04.032

Kragel, J. E., Vanhaerents, S., Templer, J. W., Schuele, S., Rosenow, J. M., Nilakantan, A. S., & Bridge, D. J. (2020). Hippocampal theta coordinates memory processing during visual exploration. *ELife*, *9*. https://doi.org/10.7554/ELIFE.52108

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

Kuchinsky, S. E., Jr., K. I. V., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2015). Task-Related Vigilance During Word Recognition in Noise for Older Adults with Hearing Loss. *Https://Doi.Org/10.1080/0361073X.2016.1108712*, *42*(1), 64–85. https://doi.org/10.1080/0361073X.2016.1108712

Kwak, C., Han, W., Lee, J., Kim, J., & Kim, S. (2018). Effect of noise and reverberation on speech recognition and listening effort for older adults. *Geriatrics and Gerontology International*, *18*(12), 1603–1608. https://doi.org/10.1111/ggi.13535

Kwon, H. B. (2010). Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses. *The Journal of Advanced Prosthodontics*, *2*(3), 71. https://doi.org/10.4047/JAP.2010.2.3.71

Lakshmi, M. S. K., Rout, A., & O'Donoghue, C. R. (2021). A systematic review and meta-analysis of digital noise reduction hearing aids in adults. In *Disability and Rehabilitation: Assistive Technology* (Vol. 16, Issue 2, pp. 120–129). Taylor and Francis Ltd. https://doi.org/10.1080/17483107.2019.1642394

Leiberg, S., Lutzenberger, W., & Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Research*, *1120*(1), 131–140. https://doi.org/10.1016/j.brainres.2006.08.066

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. In *Ear and Hearing* (pp. 37 Suppl 1, 77S–84S). https://doi.org/10.1097/AUD.0000000000000304

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M., Rosenkranz, K., & Mandic, D. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE Pulse*, *3*(6), 32–42. https://doi.org/10.1109/MPUL.2012.2216717

Luck, S. J. (2005). An Introduction to Event-Related Potentials and Their Neural Origins. *MIT Press*. https://doi.org/10.1007/s10409-008-0217-3

Luck, S. J. (2014). An Introduction to the Event-Related Potential Technique, second edition. In *The MIT Press*. https://doi.org/10.1118/1.4736938

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., & Ng, E. H. N. (2016). Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *Ear and Hearing*, *37 Suppl 1*, 145S-54S. https://doi.org/10.1097/AUD.0000000000000294

Luo, H., Husain, F. T., Horwitz, B., & Poeppel, D. (2005). Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study. *NeuroImage*, *28*(1), 59–71. https://doi.org/10.1016/j.neuroimage.2005.05.040

Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, *37*, 118S-125S. https://doi.org/10.1097/AUD.0000000000000305

Mackersie, C. L., Macphee, I. X., & Heldt, E. W. (2015). Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear and Hearing*, *36*(1), 145–154. https://doi.org/10.1097/AUD.0000000000000091

MacPherson, A., & Akeroyd, M. A. (2013). The glasgow Monitoring of Uninterrupted Speech Task (GMUST): A naturalistic measure of speech intelligibility in noise. *Proceedings of Meetings on Acoustics*. https://doi.org/10.1121/1.4799865

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*(4), 283–293. https://doi.org/10.1016/0013-4694(93)90110-H

Manza, P., Hau, C. L. V., & Leung, H. C. (2014). Alpha power gates relevant information during working memory updating. *Journal of Neuroscience*, *34*(17), 5998–6002. https://doi.org/10.1523/JNEUROSCI.4641-13.2014

Marsella, P., Scorpecci, A., Cartocci, G., Giannantonio, S., Maglione, A. G., Venuti, I., Brizi, A., & Babiloni, F. (2017). EEG activity as an objective measure of cognitive load during effortful listening: A study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss. *International Journal of Pediatric Otorhinolaryngology*, *99*, 1–7. https://doi.org/10.1016/j.ijporl.2017.05.006

Mathers, C., Smith, A., & Concha, M. (2001). Global burden of hearing loss in the year 2000. *Global Burden of Disease*. https://doi.org/10.1038/nature02612.1.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology*, *53*(7), 433–440. https://doi.org/10.3109/14992027.2014.890296

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., Miles, K., & Graham, P. L. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.00745

Mierau, A., Klimesch, W., & Lefebvre, J. (2017). State-dependent alpha peak frequency shifts: Experimental evidence, potential mechanisms and functional implications. *Neuroscience*, *360*, 146–154. https://doi.org/10.1016/J.NEUROSCIENCE.2017.07.037

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., & Kidmose, P. (2015). EEG recorded from the ear: Characterizing the Ear-EEG Method. *Frontiers in Neuroscience*, *9*, 438. https://doi.org/10.3389/fnins.2015.00438

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

Mirkovic, B., Bleichner, M. G., De Vos, M., & Debener, S. (2016). Target Speaker Detection with Concealed EEG Around the Ear. *Frontiers in Neuroscience*, *0*(JUL), 349. https://doi.org/10.3389/FNINS.2016.00349

Miyake, A., & Shah, P. (1999). Models of Working Memory. In *Models of Working Memory*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Naylor, G., Koelewijn, T., Zekveld, A. A., & Kramer, S. E. (2018). The Application of Pupillometry in Hearing Science to Assess Listening Effort. *Trends in Hearing*, *22*, 1–3. https://doi.org/10.1177/2331216518799437

Ng, E. H. N., Rudner, M., Lunner, T., & Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000080

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, *50*(3), 202–208. https://doi.org/10.3109/14992027.2010.524254

Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Nyberg, L., & Eriksson, J. (2016). Working Memory: Maintenance, Updating, and the Realization of Intentions. *Cold Spring Harbor Perspectives in Biology*, *8*(2). https://doi.org/10.1101/CSHPERSPECT.A021816

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

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(11), 2466–2477. https://doi.org/10.1093/cercor/bhr325

Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network. *Journal of Neuroscience*, *32*(36), 12376–12383. https://doi.org/10.1523/jneurosci.4908-11.2012

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90–99. https://doi.org/10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68–79. https://doi.org/10.1016/j.heares.2017.05.012

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, *27*(2), 341–356. https://doi.org/10.1016/j.neuroimage.2005.04.014

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. *Trends in Neurosciences*, *30*(4), 150–158. https://doi.org/10.1016/j.tins.2007.02.001

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, *0*(8 JUL), 253. https://doi.org/10.3389/FNINS.2014.00253

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Frontiers in Psychology*, *3*, 606. https://doi.org/10.3389/fpsyg.2012.00606

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

Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., & Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Frontiers in Psychology*, *6*, 177. https://doi.org/10.3389/fpsyg.2015.00177

Pfurtscheller, G. (2001). Functional brain imaging based on ERD/ERS. *Vision Research*. https://doi.org/10.1016/S0042-6989(00)00235-2

Pfurtscheller, G., Stancák, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band - An electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, *24*(1–2), 39–46. https://doi.org/10.1016/S0167-8760(96)00066-9

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing*, *37*(1), 1–13. https://doi.org/10.1097/AUD.0000000000000222

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

Picou, E. M., & Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *International Journal of Audiology*, *57*(6), 457–467. https://doi.org/10.1080/14992027.2018.1431696

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing*, *34*(5). https://doi.org/10.1097/AUD.0b013e31827f0431

Pinal, D., Zurrón, M., & Díaz, F. (2014). Effects of load and maintenance duration on the time course of information encoding and retrieval in working memory: From perceptual analysis to post-categorization processes. *Frontiers in Human Neuroscience*, *8*, 165. https://doi.org/10.3389/fnhum.2014.00165

Plain, B., Richter, M., Zekveld, A. A., Lunner, T., Bhuiyan, T., & Kramer, S. E. (2020). Investigating the Influences of Task Demand and Reward on Cardiac Pre-Ejection Period Reactivity During a Speech-in-Noise Task. *Ear & Hearing*. https://doi.org/10.1097/aud.0000000000000971

Plomp, R., & Mimpen, A. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333–1342. https://doi.org/10.1121/1.383554

Prodi, N., & Visentin, C. (2019). *Impact of Background Noise Fluctuation and Reverberation on Response Time in a Speech Reception Task*. *62*(11), 4179–4195. https://doi.org/10.1044/2019_JSLHR-H-19-0180

Ratnam, R., Jones, D. L., Wheeler, B. C., O'Brien, W. D., Lansing, C. R., & Feng, A. S. (2003). Blind estimation of reverberation time. *The Journal of the Acoustical Society of America*, *114*(5), 2877. https://doi.org/10.1121/1.1616578

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

Rondina, R., Olsen, R. K., Li, L., Meltzer, J. A., & Ryan, J. D. (2019). Age-related changes to oscillatory dynamics during maintenance and retrieval in a relational memory task. *PLoS ONE*, *14*(2). https://doi.org/10.1371/journal.pone.0211851

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Rosemann, S., & Thiel, C. M. (2020). Neural Signatures of Working Memory in Age-related Hearing Loss. *Neuroscience*, *429*, 134–142. https://doi.org/10.1016/J.NEUROSCIENCE.2019.12.046

Rossing, T. D. (2007). Springer Handbook of Acoustics. In *Springer*. Springer New York. https://doi.org/10.1007/978-0-387-30425-0

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., & Lunner, T. (2020). An exploratory Study of EEG Alpha Oscillation and Pupil Dilation in

Hearing-Aid Users During Effortful listening to Continuous Speech. *PLOS ONE*, *15*(7), e0235782. https://doi.org/10.1371/journal.pone.0235782

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & van Doornen, L. J. P. (1990). Methodological Guidelines for Impedance Cardiography. *Psychophysiology*, *27*(1), 1–23. https://doi.org/10.1111/j.1469-8986.1990.tb02171.x

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

Speaks, C., Parker, B., Harris, C., & Kuhl, P. (1972). Intelligibility of connected discourse. *Journal of Speech and Hearing Research*. https://doi.org/10.1044/jshr.1503.590

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective and Behavioral Neuroscience*, *17*(4), 809–825. https://doi.org/10.3758/s13415-017-0513-0

Tuladhar, A. M., Ter Huurne, N., Schoffelen, J. M., Maris, E., Oostenveld, R., & Jensen, O. (2007). Parieto-occipital sources account for the increase in alpha activity with working memory load. *Human Brain Mapping*, *28*(8), 785–792. https://doi.org/10.1002/hbm.20306

van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of Speech Clarity on Recognition Memory for Spoken Sentences. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0043753

Wang, Yang, Kramer, S. E., Wendt, D., Naylor, G., Lunner, T., & Zekveld, A. A. (2018). *The Pupil Dilation Response During Speech Perception in Dark and Light : The Involvement of the Parasympathetic Nervous System in Listening Effort*. *22*, 1–11. https://doi.org/10.1177/2331216518816603

Wang, Yang, Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). *Relations Between Self-Reported Daily-Life Fatigue , Hearing Status , and Pupil Dilation During a Speech Perception in Noise Task*. 573–582.

Wang, Yuanjia, & Chen, H. (2012). On Testing an Unspecified Function Through a Linear Mixed Effects Model with Multiple Variance Components. *Biometrics*, *68*(4), 1113–1125. https://doi.org/10.1111/j.1541-0420.2012.01790.x

Ward, C. M., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2016). Effects of Age, Acoustic Challenge, and Verbal Working Memory on Recall of Narrative Speech. *Experimental Aging Research*. https://doi.org/10.1080/0361073X.2016.1108785

Weisz, N., Moratti, S., Meinzer, M., Dohrmann, K., & Elbert, T. (2005). Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography. *PLoS Medicine*, *2*(6), e153. https://doi.org/10.1371/journal.pmed.0020153

Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: A perspective from non-invasive neuroelectrophysiology. *Hearing Research*, *307*, 16–28. https://doi.org/10.1016/j.heares.2013.07.009

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, *369*, 67–78. https://doi.org/10.1016/j.heares.2018.05.006

Wilsch, A., & Obleser, J. (2016). What works in auditory working memory? A neural oscillations perspective. *Brain Research*, *1640*(Pt B), 193–207. https://doi.org/10.1016/j.brainres.2015.10.054

Wingfield, A. (2016). Evolution of Models of Working Memory and Cognitive Resources. *Ear and Hearing*, *37 Suppl 1*, 35S-43S. https://doi.org/10.1097/AUD.0000000000000310

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology*, *17*(7), 487–497.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*, 1–32. https://doi.org/10.1177/2331216518800869

Wisniewski, M. G. (2017). Indices of Effortful Listening Can Be Mined from Existing Electroencephalographic Data. *Ear and Hearing*, *38*(1), e69–e73. https://doi.org/10.1097/AUD.0000000000000354

Wisniewski, M. G., Iyer, N., Thompson, E. R., & Simpson, B. D. (2018). Sustained frontal midline theta enhancements during effortful listening track working memory demands. *Hearing Research*, *358*, 37–41. https://doi.org/10.1016/j.heares.2017.11.009

Wisniewski, M. G., Thompson, E. R., & Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology*, *54*(12), 1916–1928. https://doi.org/10.1111/psyp.12968

Wisniewski, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., & Sullivan, S. C. (2015). Frontal midline θ power as an index of listening effort. *NeuroReport*, *26*(2), 94–99. https://doi.org/10.1097/WNR.0000000000000306

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience*, *35*(4), 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015

Wöstmann, M., Lim, S. J., & Obleser, J. (2017). The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. *Cerebral Cortex*, *27*(6), 3307–3317. https://doi.org/10.1093/cercor/bhx074

Wright, R. A. (2008). Refining the Prediction of Effort: Brehm's Distinction between

Potential Motivation and Motivation Intensity. *Social and Personality Psychology Compass*, *2*(2), 682–701. https://doi.org/10.1111/j.1751-9004.2008.00093.x

Wu, Y.-H., Aksan, N., Rizzo, M., Stangl, E., Zhang, X., & Bentler, R. (2014). Measuring listening effort: Driving simulator versus simple dual-task paradigm. *Ear and Hearing*, *35*(6), 623–632. https://doi.org/10.1097/AUD.0000000000000079

Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort. *Ear and Hearing*, *37*(6), 660–670. https://doi.org/10.1097/AUD.0000000000000335

Xia, J., Xu, B., Pentony, S., Xu, J., & Swaminathan, J. (2018). Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *143*(3), 1523–1533. https://doi.org/10.1121/1.5026788

Yoho, S. E., Borrie, S. A., Barrett, T. S., & Whittaker, D. B. (2019). Are there sex effects for speech intelligibility in American English? Examining the influence of talker, listener, and methodology. *Attention, Perception, and Psychophysics*, *81*(2), 558–570. https://doi.org/10.3758/S13414-018-1635-3/FIGURES/3

Zekveld, A. A., DJ, H., IS, J., NJ, V., & Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, *101*, 76–86. https://doi.org/10.1016/J.NEUROIMAGE.2014.06.069

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 2331216518777174–2331216518777174. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284. https://doi.org/10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*, 498–510. https://doi.org/10.1097/AUD.0b013e31820512bb

Zhao, S., Bury, G., Milne, A., & Chait, M. (2019). Pupillometry as an Objective Measure of Sustained Attention in Young and Older Listeners. *Trends in Hearing*, *23*, 1–21. https://doi.org/10.1177/2331216519887815

Zuchowicz, U., Wozniak-Kwasniewska, A., Szekely, D., Olejarczyk, E., & David, O. (2019). EEG phase synchronization in persons with depression subjected to transcranial magnetic stimulation. *Frontiers in Neuroscience*, *13*(JAN), 1037. https://doi.org/10.3389/fnins.2018.01037