Analysis of the EEG-measured speech-FFRs of Musicians and Non-Musicians

Bachelorarbeit aus der Physik

Vorgelegt von

Melanie Wolf

17.04.2024

Lehrstuhl für Theoretische Physik I

Friedrich-Alexander-Universität Erlangen-Nürnberg

Betreuer: PD Dr. Michael Schmiedeberg

Arbeit angefertigt am Lehrstuhl für Sensorische Neurotechnologie bei Prof. Dr. Tobias Reichenbach (Department Artificial Intelligence in Biomedical Engineering, Technische Fakultät, FAU)

Abstract

For this bachelor thesis, the aim was to analyse EEG-measured speech-FFRs of musicians and non-musicians. During the EEG measurements, the participants were simultaneously presented two audiobooks, to determine whether there would be differences between the neural responses when a certain narrator was attended or ignored. Furthermore, it was investigated whether the two groups of musicians and non-musicians exhibit different brain responses in general, and in regards to the attentional modulation.

The provided EEG data was used together with two acoustic features, which were derived from the audiobooks, to compute Temporal Response Functions (TRFs). These Temporal Response Functions (TRFs) describe the relationship between the auditory stimulus and the EEG recordings. Then, the magnitudes of the Temporal Response Functions (TRFs) were determined and statistical tests were performed on the results concerning the different attentional modes (attended and ignored) and the two participant groups (musicians and non-musicians).

It was found, that on the population average level of all considered participants, the neural responses to both acoustic features showed a tendency to be stronger in the ignored mode than in the attended mode, which was statistically backed for certain latencies after the auditory stimulation.

At the considered latencies after the auditory stimulus, no significant differences could be determined between the strengths of the brain responses of the musicians and the non-musicians. This applies to the responses to both used acoustic features.

Within the group of the non-musicians, the differences between the responses in the two attentional modes were insignificant for most of the considered latencies after the auditory stimulus. Within the group of the musicians, the differences between the responses when attending or ignoring the narrator seemed greater. At the certain latencies, where significant differences were found, the ignored response was again stronger than the attended response. This was observed for both acoustic features, respectively.

Lastly, two different approaches to the calculation of the magnitudes of the TRFs were tried, by changing the order of the averaging steps in the computation process. The approaches led to two slightly different outcomes.

Contents

List of Figures

List of Tables

List of Abbreviations

FFR Frequency-Following Response
Auditory evoked potential reflecting the brain's encoding of auditory stimuli.

EEG Electroencephalography
Measurement of the potential field on the scull, which results from neural activity in the brain.

MEG Magnetoencephalography
Measurement the magnetic field on the scull, which results from the currents caused by the neural activity in the brain.

TRF Temporal Response Function
Describes the relationship between an auditory stimulus and the corresponding resulting brain response.

1 Introduction

Humans have the ability to use their auditory skills to communicate with others, to find orientation in their surroundings and to create and listen to art in the form of music. Depending on one's upbringing, interests, genetic makeup and potential neurological or hearing impairments, these auditory skills can vary from one person to another, and therefore alter the difficulties someone has in the above mentioned situations. In order to approach such difficulties, it is important to get a full understanding of the functionality of the different stages of general auditory and speech processing. Even though there is insightful work on this topic, it still has not been fully understood so far, making it an exciting field of interdisciplinary research.

For this bachelor thesis, the aim was to analyse the Frequency-Following Response (FFR) to speech in musicians and non-musicians. A FFR is the auditory evoked potential reflecting the brain's successful encoding of auditory stimuli. In this case, the presented stimulus was non-repetitive continuous speech, which is why the response can be referred to as speech-FFR. The most common methods to measure FFRs are Electroencephalography (EEG) and Magnetoencephalography (MEG). They are both non-invasive measuring techniques that capture the neural responses in the brain. These detected neural responses can be of subcortical [12, 11] or cortical [6, 28, 26] origin.

Before researching continuous speech processing, there were studies conducted to analyse the brain responses to repetitive and non-repetitive short stimuli. For instance, Coffey et al. [6] investigated the cortical responses to the periodic presentation of the speech syllable (/da/) with MEG measurements. Furthermore, Bidelman et al. [2] focused on the EEG-measured responses to clicks and sustained single-pitch stimuli. More recently, the presented auditory stimuli got longer and more similar to real-life speech, by using audio books [28] or speech samples [12]. This also enables a more precise depiction of the timing of the neural responses than with the short stimuli.

Additionally, the settings, in which the stimuli were presented got more realistic. In real life, we are often confronted with multiple sound sources, for example when walking along a busy road or attending a crowded birthday party. In these scenarios, we have to recognise and concentrate on the source we deem most important, such as the ringing of a bicycle bell or the friend we want to have a conversation with. To simulate such settings, studies started implementing a competing-speaker scenario. In doing so, a recent study showed the attentional modulation of the cortical contribution to the speech-FFR, which was measured with MEG [28]. Another fairly recent study by Etard et al. [11] demonstrated how the attentional modulation of the EEG-measured brainstem response to speech can be used to decode the attentional focus of a listener.

One aspect, that is thought to influence one's auditory skill set, is musical training. It has been shown, that musical training has a positive effect on brain plasticity [16]. More studies investigated the impact of musical training, especially in early stages of life, on different characteristics. For instance, Pantev et al. [22] found an increased auditory cortical representation in musicians compared to a control group of subjects that had never played an instrument. The musicians showed stronger MEG-measured responses to piano tones than the control subjects. Moreover, a study by Strait et al. [31] concluded, that musical training creates enhancement and efficiency of subcortical responses to emotionally charged vocal expressions. Additionally, it was also found that early musical practice leads to a better performance in tasks involving sensorimotor integration and timing [34].

Riegel [26] recently combined the issues of the effects of musical training and focus of attention on the speech-FFR. The MEG-measured responses showed significant differences in the cortical contributions to the speech-FFRs between musicians and non-musicians. However, after evaluating the data of even more subjects than in [26], these significant differences are not present anymore. These newer results will be published in the future.

To further investigate those findings of the research group, this bachelor thesis evaluates the EEG data that was simultaneously recorded to the MEG data used in [26] and [28], considering the influence of attention and musical training on the speech-FFRs. Therefore, this work is structured as follows. First, the general theoretical background is illuminated by explaining how sound is defined in a physical and physiological way, by taking a look at the human auditory system and by outlining the functionality of the EEG. Then, the methods which were used for the acquisition of the data and some of the processing software that was provided by the research group are depicted. Afterwards, a detailed presentation of the methods I used for the further treatment of the data, together with some difficulties I encountered, is given. Following this, the results of the analysis of the EEG data concerning the attentional modulation and the influence of musical training are shown. These results are then discussed in the next chapter. Lastly, the work of this bachelor thesis is concluded and an outlook on further improvements and questions is given.

2 Background

This work covers the effects of auditory attention and the influence of musical training on brain signals. The following sections intend to provide some insight on the characteristics of sound, an overview of the human auditory system and the principles behind the measuring technique employed in this study, the EEG.

2.1 Characteristics of sound

In the Oxford dictionary, sound is defined as "The sensation produced in the organs of hearing when the surrounding air is set in vibration in such a way as to affect these; also, that which is or may be heard; the external object of audition, or the property of bodies by which this is produced. Hence also, pressure waves that differ from audible sound only in being of a lower or a higher frequency." [30]. This definition gives a mixture of the physical and physiological meaning of sound, which will be further explained in the following.

2.1.1 Sound in physics

In physics, sound is an ensemble of mechanical waves, that propagate through a medium, such as a gas, liquid or solid. A mechanical wave originates from a vibrating part of the medium, which is coupled with its neighbouring areas. Due to this coupling, the oscillation energy gets transferred to and excites the surrounding area. The distance such a wave can cover depends on the medium and its properties.

The spectrum of physical sound can be divided into four parts. Waves with frequencies smaller than $16\,\mathrm{Hz}$ are called infrasound, waves with frequencies larger than $16\,\mathrm{kHz}$ are called ultrasound and waves with frequencies even higher than $10\,\mathrm{MHz}$ are called hypersound. The sound waves that can actually be detected by the human ear have frequencies between $16\,\mathrm{Hz}$ and $16\,\mathrm{kHz}$ [8].

Audible sound waves can be categorized even further. A harmonic oscillation with a constant amplitude is called a tone. Its frequency determines the pitch. A superposition of different tones is then called a sound [8].

2.1.2 Sound and speech in human physiology

The definition of sound in a more physiological sense starts, where the physical description ends. The above mentioned sound waves cause a change in air pressure, which can be detected by the human ear. The ear is one of the human sensory organs and sound is therefore an important part of human communication and orientation in ones surroundings. Sounds can be broken down into pitch, timing and timbre [16]. The pitch is the perceived frequency of a sound and the timing marks specific events in a sound signal, such as the onset. Timbre depicts the quality of a sound by combining temporal and spectral attributes of the sound signal [16].

Speech is the type of sound someone produces, in order to communicate with someone else. It is a composition of ones breathing, larynx, pharynx and mouth [3]. Fundamental tones are the result of the vocal folds setting the exhaled air into vibration. The amplitude and frequency of the emerging sound wave depend on the air pressure of the exhale, as well as the properties of the vocal folds. Further modulation of the sound wave (enunciation) follows in the pharynx and mouth. Sufficient auditory abilities are highly important for the development of speech [3].

2.2 The human auditory system

To investigate the effects of auditory attention and musical training on the brain signals caused by sound, it is necessary to understand the functionality of the human auditory system, i.e., how a signal is processed within its auditory pathway. Therefore, the following two subsections give an overview of the peripheral and central auditory system.

2.2.1 The peripheral auditory system

The ear is the peripheral auditory system of the human body. As shown in Figure 2.1, it consists of three areas, the outer, middle and inner ear. The outer ear is composed of the pinna and the ear canal. The eardrum connects the outer ear to the middle ear and vibrates as a result of the change in air pressure mentioned in subsection 2.1.2. The middle ear is made up of the malleus, the incus and the stapes. They are three tiny bones, called the ossicles, which amplify the vibrations and transfer them, through the oval window, to the inner ear [25].

The inner ear consists of the semicircular canals and the cochlea. The cochlea has a snail-like shape and is filled with a fluid, as well as the so-called basilar membrane. The fluid transfers the incoming vibrations to the basilar membrane, which then starts to oscillate itself. Atop the membrane lie hair cells, which act as sensory cells for the incoming sound waves. As the cochlea gets narrower toward its center, the frequencies that are detected get lower. Therefore, at the wider end of the cochlea, the hair cells detect higher-pitched sounds, whereas the hair cells that lie more towards the center detect lower-pitched sounds. Due to the vibration of the basilar membrane, the hair cells move as

well. This movement causes the stereocilia, microscopic appendices that are located on top of the hair cells, to press against the surrounding walls and bend. This bending leads to channels at the tips of the stereocilia to open up, through which chemicals can enter the cells and produce an electrical signal. This electrical signal then travels through the auditory nerve to the brain, where the further processing and comprehension of a sound takes place [7].

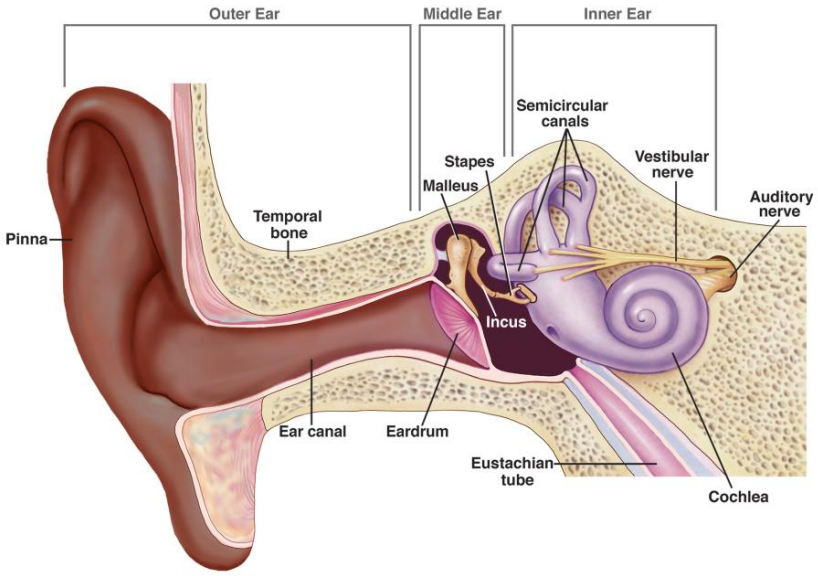


Figure 2.1: Schematic depiction of the peripheral human auditory system. It consists of the outer, middle and inner ear. Sound travels through the ear canal in the outer ear and sets the eardrum into vibration. The malleus, incus and stapes amplify these vibrations and transfer them through the oval window to the inner ear. In the cochlea, the vibrations cause electrical signals which then travel through the auditory nerve to the brain. The image was taken from [7].

2.2.2 The central auditory system

After a sound was received by and travelled through the ear as described in subsection 2.2.1, the auditory nerve transfers the resulting electrical signal to the brain, where the contained information gets processed. The auditory pathways run through the brainstem, the midbrain and the interbrain to the auditory cortex [3]. The involved nuclei are depicted in Figure 2.2. The figure shows the afferent connections, which transport the sensory input to the brain, between the nuclei starting from the right ear. The midbrain, specifically the Colliculus inferior, and the auditory cortex combine the information gathered in the lower parts of the auditory pathways. In the cortex specifically, the auditory information is connected to the information received from other sensory organs [3].

Some of the most important elements of sound that get processed by the central auditory pathways are intensity, attenuation, spatial location and frequency [25]. These different attributes require different types of auditory neurons or respective reactions of the neurons to encode the information.

For the processing of a sounds intensity and attenuation, neurons fire at different rates. A louder sound results in a higher firing rate of the neurons. There are also specialized neurons whose firing rates peak within different ranges of sound intensities [25]. The analysis of frequency-dependant amplitudes and phases of the signal already starts at the Nucleus cochlearis in the brainstem [3].

For the processing of a sounds spatial location, the central auditory pathways are 'monaural' in the lower parts and 'binaural' from the superior olivary complex (Nucl. olivaris sup.) upwards [25, 3]. This means, that the inputs from both ears travel to the superior olivary complex within separate pathways, which then merge into one mutual pathway. If the origin of a sound is, for example, on the right side of a persons head, the superior olivary complex receives the input from the right auditory nerve faster than the input from the left ear, resulting in a temporal delay [25]. Additionally, the difference in intensities from the ear that is facing the sound source versus the ear that is averted from the sound source is compared in the nucleus olivaris superior [3].

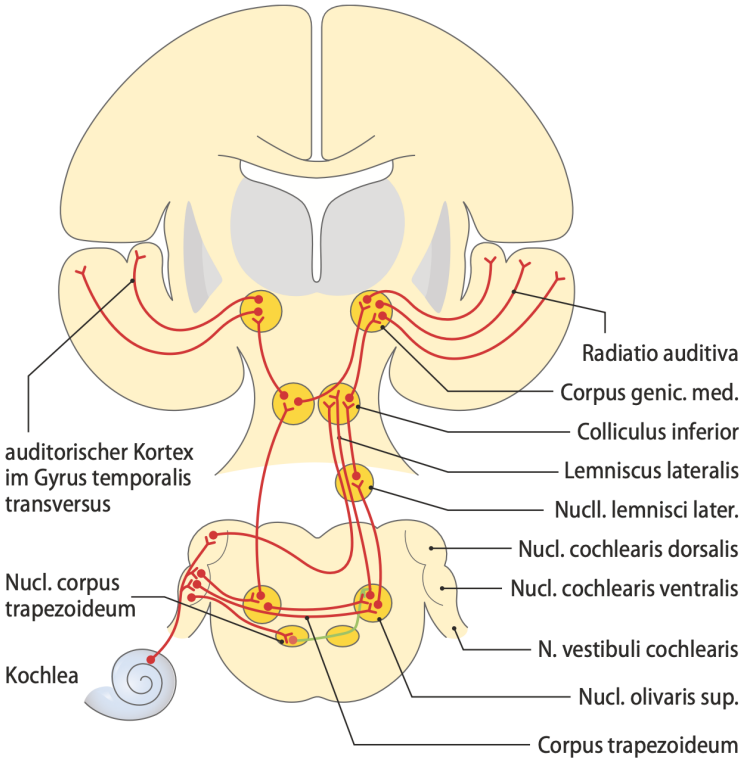


Figure 2.2: Depiction of the afferent auditory pathways starting from the right ear. The pathways run through the brainstem, the midbrain and the interbrain to the auditory cortex [3].

2.3 Electroencephalography

EEG is a non-invasive measuring technique to record electric activity in the brain. It is widely used for diagnostic purposes in medicine and has also gained popularity as a research method for brain activity in neuroscience.

For an EEG measurement, electrodes made of conductive materials are placed on the scull via an electrode cap. There are two types of electrodes, 'active' and 'passive' ones. The passive electrodes are connected to an amplifier by cable, while the active electrodes have inbuilt preamplifiers [19]. This leads to the active electrodes being less sensitive to noise but also not suitable for e.g. combined EEG and MEG measurements, since the electromagnetic fields of the preamplifiers would disturb the MEG recording. The EEG caps can differ in the number of channels they provide. To enhance the electric conductivity, electrolytic gels are often used between the electrodes and the scull [1].

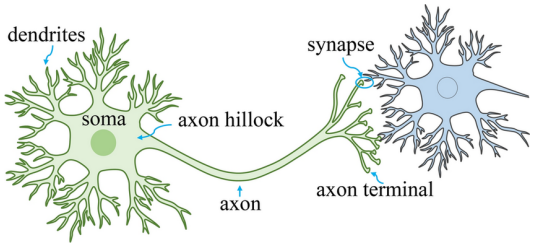


Figure 2.3: Schematic depiction of two connected neurons. At the synapses, the axon terminals of the left (green) neuron release neurotransmitters, which are received by the right (blue) neuron [18].

The electrodes detect postsynaptic potentials which result from the release of neurotransmitters at the axon terminal of a neuron. Figure 2.3 shows the schematic structure of two connected neurons. The neurotransmitter is received by a dendrite of the connected neuron, leading e.g. to an intracellular current source and an extracellular current sink. As a result, there will be an intracellular current sink and an extracellular current source at the soma of the second neuron [1].

Since the EEG electrodes are placed on top of the scull, there are several layers, like cerebral fluid and hair, between the electrodes and the area of neural activity. Therefore, the above described process has to happen simultaneously for a larger number of neurons, for the activity to be strong enough to be captured. The neurons in the cortex lie parallel to each other and orthogonal to the surface of the cortex [32]. For a synchronized synaptic input, the aligned neurons all show the same polarization in extracellular and reversed intracellular dipoles. The resulting opposing currents lead to a configuration of sinks and sources in the different layers of the cortex, which furthermore lead to electrical fields that can be measured as potential fields on the surface of the scull [32].

While the EEG is sensitive to those cortical currents closer to the electrodes, it can also sense deep

currents, such as auditory brainstem responses. Since these currents occur in the brainstem, they are measured earlier, at smaller latencies, than the cortical contributions to the brain response [24]. The latencies generally occur, because the input that is captured by the ear does not reach the brain instantaneously, but with a temporal delay. The auditory nerve, that transfers the electrical signal from the ear to the brain, travels through the subcortex to the cortex, hence the earlier responses in the subcortical region of the brain. A more detailed description of the human auditory pathways is presented in section 2.2. The subcortical responses can generally be measured more easily with EEG compared to, for instance, MEG. MEG measures the magnetic fields that arise from the electrical currents in the brain and is another popular method to capture neural activity.

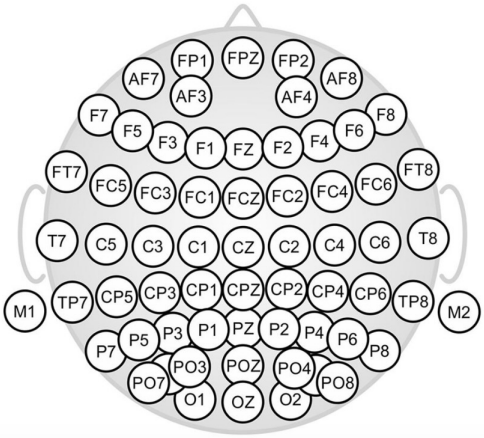


Figure 2.4: Exemplary placement of 64 electrodes with the extended 10-20 system. The image was taken from [13]. The electrodes are labeled, with the letters corresponding to the cortical location (frontal, temporal, parietal, occipital, central and mixtures of those) and the numbers indicating the laterality (odd numbers left, even numbers right, 'z' zero) [19].

A scheme that is commonly used for EEG measurements is the so-called 10-20 system. The number of used electrodes can vary for different setups. The areas of the brain beneath each electrode are proportional to the size of the head. Therefore, the observed region of the brain beneath the respective electrode stays the same for each person. To correctly align the electrodes, four reference points are used: the nasion (at the nose), the inion (at the back of the scull) and the preauricular points (at the ears), one on each side of the head [21]. With the extended 10-20 system, which was used for the measurements of the data analysed in this study, the number of electrodes is increased to, in this case, 64, allowing for more measured data in general and a better understanding of the spatial origin of a signal. The electrodes are labeled, with the letters corresponding to the cortical location (frontal, temporal, parietal, occipital, central and mixtures of those) and the numbers indicating the laterality (odd numbers left, even numbers right, 'z' zero) [19]. An example for a theoretical placement of the extended 10-20 system with 64 electrodes is depicted in Figure 2.4.

3 Methods used for provided data and software

The following chapter outlines the steps in the acquisition of the raw EEG data, as well as the methods behind the software that was already provided to me by the research group.

3.1 Acquisition of data

The EEG data was recorded along with MEG data by Alina Schüller and Jasmin Riegel for their studies [26, 28]. However, the EEG data has not been analysed or worked with previous to this bachelor thesis.

3.1.1 Experimental procedure

Two audiobooks, narrated by two male speakers, were simultaneously presented to the participants, while they had to focus their attention on only one of the two speech signals. The two speakers could be differentiated by the pitch of their voices. The first speaker had, on average, a slightly lower pitch, with a fundamental frequency range of approximately 70 Hz to 120 Hz, while the second speaker had a fundamental frequency range of approximately 100 Hz to 150 Hz. Therefore, the first narrator is referred to as the lower pitch speaker and the second narrator as the higher pitch speaker [26, 28].

In total, four different audiobooks were used, consisting of the two story-audiobooks "Frau Ella" by Florian Beckerhoff and "Den Hund Überleben" by Stefan Hornbach, as well as the two noise-audiobooks "Darum" by Daniel Glattauer and "Looking for hope" by Colleen Hoover (translated to German by Katarina Ganslandt). The story-audiobooks are the ones on which the participants should focus, while being distracted by the noise-audiobooks. "Frau Ella" and "Darum" were narrated by Peter Jordan, the lower pitch speaker, while "Den Hund überleben" and "Looking for hope" were narrated by Pascal Houdus, the higher pitch speaker. All audiobooks were published by *Hörbuch Hamburg* [26, 28].

The participants concurrently listened to the first story-audiobook and the second noise-audiobook, having to focus their attention on the story-audiobook. After the first chapter, the participants had to attend the second story-audiobook, while ignoring the first noise-audiobook. Then they had to switch their attention to the first story-audiobook again, and so on. The timing of switching ones attention was determined by the chapter lengths of the story-audiobooks. For the corresponding ignored parts random segments of the respective noise-audiobooks were chosen and cut to fit the duration of the chapters of the story-audiobooks. The participants listened to ten chapters in total, resulting in approximately 40 minutes of measured data. To signal which narrator the participants should attend,

the audio files of the story-audiobooks were started approximately five seconds earlier than those of the noise-audiobooks. A more detailed explanation can be found in [26] and [28].

Although the data was recorded for both speakers, only the data in which the lower pitch speaker was attended and ignored is further considered in this bachelor thesis. The previous work of Schüller et al. [28] on the corresponding MEG data of 16 of the 50 subjects (and additional six subjects that were not considered here) revealed a significantly higher neural response to the lower pitch speaker than to the higher pitch speaker.

3.1.2 Participants

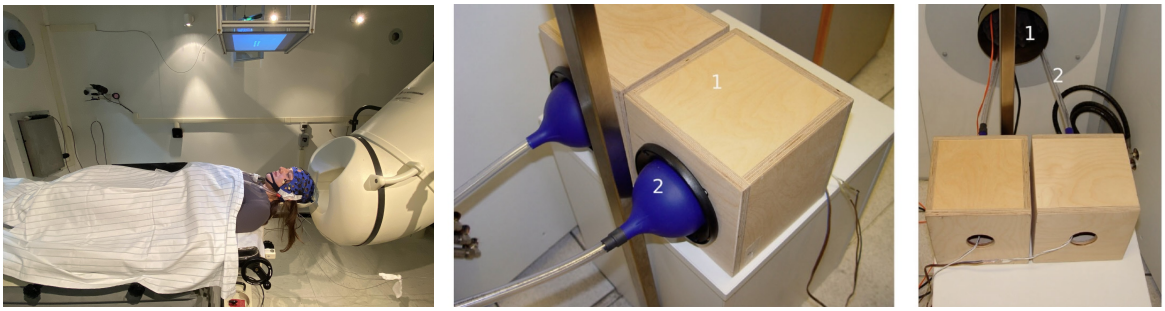
For this bachelor thesis, 50 participants (25 female), aged 18 to 30 years, were evaluated in total. 16 of them participated in the study of Schüller et al.[28] and Riegel [26], 27 of them in the work of Riegel [26] and the data of the remaining 7 of them was only recently measured to enlarge the available datasets for future analysis within the research group. All 50 subjects were native German speakers, right handed and had no history of hearing impairments or neurological diseases. The ethics board of the University Hospital Erlangen granted permission for this study (registration no. 22-361-S) [26, 28].

The participants were divided into groups of musicians and non-musicians. The criteria for this discrimination are further explained in section 3.2. In total, 18 subjects could be categorized as musicians and 23 as non-musicians. As previously mentioned, out of the 50 subjects, 16 were participants in the study of Schüller et al. [28] and retrospectively questioned on their musical background. Two of them turned out to be non-musicians and five of them to be musicians. The remaining nine subjects didn't fit into either of those two groups, but were still considered for those parts of the analysis in this work that did not involve the influence of musical training.

3.1.3 Experimental setup

Since MEG data was measured simultaneously to the EEG data, the participants were in a magnetically shielded room during the recordings. The setup for the speech presentation and the MEG measurement was already provided through a study by Schilling et al. [27].

For the EEG measurement, a MEG suitable passive 64-electrode system from ANT (TMSI REFA-8) was used, which was powered by an ASA-Lab battery box. The participants put on an EEG cap (see Figure 3.1a) with 64 pre-installed Ag/AgCl sintered MCScap-T electrodes (GVB geliMED SLEEP), additionally an electrode contact gel was applied beneath each electrode to provide high electrical conductivity. The electrode cap was designed specifically for combined EEG and MEG measurements. Additionally, two electrodes were affixed beneath and above the right eye to record electrooculography data, as well as one on the chest to record electrocardiography data. This was



(a) (b)

Figure 3.1: (a) The equipment for the EEG measurement was placed inside of the chamber, on the other side of the participant in this picture. The shielded cables were then led through the wall to the PC on the outside. (b) left: On the outside of the MEG cabin, two loudspeakers (1) are connected to two funnels (2) which are each connected to silicone tubes. (b) right: The silicone tubes (2) enter the chamber through a small hole (1) to transmit the sound. The images in (b) were taken from [27].

implemented to enable artifact exclusion due to eye movement, measured by the electrooculogram, and heart activity, recorded by the electrocardiogram, in the preprocessing of the EEG data later on (see section 3.3). The used devices were all placed inside the MEG chamber and connected with shielded cables. The cables were then led through a hole in the wall to a separate computer, which was placed on the outside of the chamber. The EEG data was recorded with a sampling frequency of $2000\,\mathrm{Hz}$, meaning every $0.0005\,\mathrm{s}$ one separate datapoint was measured.

Additional to the PC for the EEG measurement, another two separate computers were used, one to record the MEG data and one for the acoustic stimulation, presenting the aforementioned audiobooks. The computers and equipment were interconnected in a way that ensured a precise temporal alignment of the audio stimulus and the measured data afterwards (see section 4.1). A more detailed depiction of these connections and of the used equipment for the audio presentation and MEG measurement can be found in [26] and [27].

The setup for the audio presentation is depicted in Figure 3.1b. The two loudspeakers, one for each ear, were positioned outside of the shielded chamber to prevent artifacts caused by the magnetic fields of the speakers. The sound was then transmitted through silicone funnels coupled with flexible tubes ($\approx 2\,\mathrm{m}$ in length, $\approx 2\,\mathrm{cm}$ inner diameter). The tubes were led through a small hole in the wall of the MEG chamber (see Figure 3.1b). Resulting from the length of the sound path from the stimulation PC to the participant's ear, a constant temporal delay of $\approx 6\,\mathrm{ms}$ was determined. This was taken into account for the temporal alignment of the measured data and the audio stimulus later on.

3.2 Definition of musicality

As this bachelor thesis analyses EEG-recorded speech-FFRs of musicians and non-musicians in a two-speaker scenario, this section explains the criteria which were used to distinguish the participants.

The participants in this bachelor thesis were divided into three disjoint groups: musicians, non-musicians and those that didn't fit into either category. As the topic of this work is strongly inspired by a previous study conducted within the research group by Riegel [26], the same requirements were chosen for the categorization of the participants. A more detailed explanation for the selection of these requirements is therefore given in [26]. A summary of the conditions to count as a musician or a non-musician is given in Table 3.1. In total, three factors were considered for the differentiation of the subjects. The first one being the age, at which the person started his or her musical training, the second one being the total number of years the person was in musical training and the third one being whether the person was still in active musical training.

The relevance of the starting age for auditory skills has been established in various studies [16, 34, 23, 22]. For example, Pantev et al. [22] found that the starting age of musical training correlates linearly with a subjects' ability to recognize piano tones. The younger the subjects started their musical training, the more enhanced their cortical responses were. For this work, the threshold for starting ones musical training was chosen to be age seven. To qualify as a musician, one has to have started their musical training before the age of seven. If that is not the case, the participant would either qualify as a non-musician, or neither of the two.

As found in e.g. [23] and [31], the total years of musical training have an impact on various aspects of auditory skills. For example, in a study by Strait et al. [31] a connection between the years of musical training and the timing of brain responses was shown. Participants that were in musical training for more than ten years showed faster responses to emotional sound cues than the non-musicians. In this bachelor thesis, the threshold for the years of musical training was chosen to be ten years for a participant to qualify as a musician. To count as a non-musician, a participant couldn't have been in musical training for more than three years.

The third criterion was the years since the participants' last training period. For this work, the boundary was set to zero. Therefore, if for a participant the period since his/her last training was more than zero years, he/she did not qualify as a musician. This means that musicians are currently actively playing an instrument, while non-muscians are currently not in musical training.

To actually count as a musician or a non-musician, a participant had to meet all of the required criteria. For instance, if someone started their musical training before the age of seven but only played for four years, they wouldn't meet the criteria for either of them and therefore qualify for the third

	Starting Age	Years of Training	Currently in Training
Musicians	≤ 7	≥ 10	yes
Non-Musicians	≥ 7	≤ 3	no

Table 3.1: Criteria of musicians and non-musicians.

group.

Considering the above mentioned criteria, out of the 50 participants, 18 qualified as musicians, 23 as non-musicians and nine as neither. Those nine subjects were still included for the analysis that did not focus on the influence of musical training.

3.3 Preprocessing of the EEG data

Before the recorded EEG data could be analysed, some preprocessing needed to be implemented. To some extent, Python scripts were therefore used, that were provided by the research group. The setup and functionality of these scripts is described in the following. Additional work and modulations that were added during my own work on the data are then explored in section 4.1. For the preprocessing, as well as parts of the eventual analysis, the MNE-Python package was used, which is introduced in [14]. It is an open source package for the analysis and visualisation of neurophysiological data, such as EEG data, and provides special functions for the work on such datasets.

Firstly, the EEG and MEG data sets of a subject were temporally aligned, which is explained in detail in section 4.1. Afterwards, the EEG channels were spatially aligned on the subject's head by creating a montage with subject-specific digitized channel positions. Then, the EEG data, which was recorded with a sampling frequency of 2000 Hz, was down-sampled to 1000 Hz and several notch filters were applied to exclude potentially captured power supply line signals. As the EEG measures potentials, a reference value of voltage is needed. To do so, a virtual electrode was used, that was created by averaging over all of the 64 available channels. The process of choosing this as the preferred reference option is depicted in section 4.1. Channels that showed artifacts, which are variations in the EEG data that are not caused by neural activity, or that did not record any signal at all were selected and interpolated with respect to their neighbouring electrodes (see section 4.1). Lastly, components in the EEG data that were found to be related to the measured heart activity (electrocardiography data) and eye movement (elecrooculography data) were excluded. The preprocessed data was then temporally aligned with the corresponding audiobook, respectively for the attended and the ignored mode of the lower pitch speaker. Therefore, for each subject, there were now two separate preprocessed and temporally aligned data files. One containing the EEG data, where the subject attended the lower pitch speaker (attended mode) and one, where the subject attended the higher pitch speaker and therefore ignored the lower pitch speaker (ignored mode). Finally, the EEG data was bandpass-

filtered in the frequency range of the lower pitch narrator, since only the neural responses to the lower pitch speaker should be investigated within this bachelor thesis. The lower edge cut-off frequency was set to $70\,\mathrm{Hz}$ and the upper edge to $120\,\mathrm{Hz}$, as suggested by the fundamental frequency histogram of the narrators' voices in [28]. The filter was a linear digital Butterworth bandpass filter of fifth order, which was applied forward and backward.

3.4 Selection and extraction of the acoustic features

To investigate the neural responses in the competing speaker scenario, two acoustic features of the presented audiobooks narrated by the lower pitch speaker were chosen for the further analysis. These two features, the fundamental waveform $f(t)$ and the envelope modulation $e(t)$ of its higher harmonics, were also the selected features in previous studies within the research group, that focused on the analysis of the corresponding MEG data [26, 28]. Thus, they were already provided by the research group. While working with the EEG data however, other acoustic features were tried as well. Their selection and extraction is explained in section 4.2, together with a discussion about their respective suitability for the analysis in this bachelor thesis.

The fundamental waveform is the waveform of the speech signal that oscillates at the fundamental frequency f_0 of the speaker's voice. f_0 was extracted from the audio files with the Probabilistic YIN algorithm [20], separately for each chapter of the audiobooks narrated by the lower pitch speaker. For the subsequent filtering, the cut-off frequencies resulted in $65\,\mathrm{Hz}$ for the lower edge and $120\,\mathrm{Hz}$ for the upper edge. A more detailed explanation of the extraction process can be found in [28].
The envelope modulation of the higher modes of the fundamental frequency has been shown to contribute significantly to the neural response as well [28, 15, 17]. To extract $e(t)$ from the original audio file, a model was used which approximates the transformation of the speech signal as it travels through the auditory system. It therefore models the first steps of the auditory pathway that a signal passes. The used Matlab code by Chi et al. [5] was translated to Python within the research group. Again, for a further description of the extraction process see [28].

3.5 Temporal Response Functions

The brain response to the speech-stimuli was decoded by employing TRFs. By using a linear forward model, the TRF describes the relationship between the auditory stimulus and the EEG recordings. The TRF can then be plotted over latency intervals, relative to the acoustic signal that caused the alteration in the EEG data. The Python script for the estimation of the TRF coefficients was already provided by the research group and previously developed by Etard et al. [11] and Kegler et al. [15]. The changes in the script and other considerations that were necessary for this bachelor thesis are explained in section 4.3.

The EEG-measured neural response $y^{(c)}(t)$ at a time t and a channel c is represented by the summation over different latencies τ of the linear combination of the acoustic features of the speech stimulus at times $t - \tau$. The acoustic features in this case consist, as previously described, of the fundamental waveform $f(t-\tau)$ and the envelope modulation of its higher modes $e(t-\tau)$. Additionally, the audio features are multiplied by their respective weights $\alpha^{(c)}(\tau)$ and $\beta^{(c)}(\tau)$ at the latency τ:

$$y^{(c)}(t) = \sum_{\tau=\tau_{min}}^{\tau_{max}} (\alpha^{(c)}(\tau) \cdot f(t-\tau) + \beta^{(c)}(\tau) \cdot e(t-\tau)) \tag{3.1}$$

Said weights are the TRFs of the particular speech feature. By estimating the weights in consideration of both acoustic features, as seen in Equation 3.1, the results are more realistic compared to the independent calculation of the weights. That is, because both acoustic features also contribute to the neural response. The feature that contributes more to the neural response shows a larger TRF amplitude [17]. The here considered latency range is $\tau_{min} = -20\,\mathrm{ms}$ to $\tau_{max} = 120\,\mathrm{ms}$ with $1\,\mathrm{ms}$ increments. The negative latencies are shown to ensure that there are no brain responses to the stimulus before the stimulus was even presented. The second boundary was chosen to be $\tau_{max} = 120\,\mathrm{ms}$, since there were no brain responses observed past this latency, but to still give a broad enough range to depict the responses above noise level at earlier latencies.

To employ the linear forward model, regularized ridge regression was used as a regularization technique, as it was the chosen method in previous studies conducted in the research group [29, 26]. The regularisation is generally necessary to avoid overfitting of the model to the training data, which would lead to the model not working as satisfactorily for new test data. Regularized ridge regression is a regularization technique that shrinks the estimated coefficients in linear models, such as α and β in Equation 3.1 [10]. The shrinkage, and therefore the coefficients, depend on a regularization parameter $\lambda \geq 0$. The bigger λ, the greater the shrinkage - or regularization - of the coefficients. By performing cross validation, a regularization parameter of $\lambda = 5$ was found to be the most adequate for all subjects.

As mentioned above, the most suitable regularization parameter was evaluated by performing five-fold cross validation on the data. It is a commonly used validation method, especially for smaller data sets, where a separate independent test set is not available. All of the available data gets split into N equally sized subsets ($N = 5$ subsets in this case). Now, one subset is used as a test set, while the remaining $N - 1$ subsets are used as training data for the model [10]. This procedure is repeated N times using a different subset as test dataset each time. For every iteration, a prediction error is estimated. By averaging over all N errors, a total prediction error can be determined for each λ, respectively, making it possible to rank the λ in terms of their suitability. The evaluation for the here chosen regularization parameter of $\lambda = 5$ is explained further in section 4.3.

4 Methods tried and used specifically in this work

This chapter illuminates the methods that were tried and used specifically for the work on the EEG data for this bachelor thesis. It depicts some of the steps of the preprocessing of the raw EEG data in greater detail. Then, the other acoustic features that were tried for the analysis, but discarded, are introduced. Afterwards, the further evaluation process of the chosen regularization parameter λ is explained, followed by the description of the alignment check, which was implemented in order to improve the temporal alignment of the EEG data and the acoustic features of the audiobooks. Lastly, the two different options for the magnitude computation that were tried in this work are shown, as well as the procedures for the statistical tests that were conducted on the results. This chapter therefore gives a detailed insight into the methods that made up the majority of my work on the EEG data for this bachelor thesis.

4.1 Preprocessing of the EEG data

In this section, the preprocessing steps of the EEG data that were only mentioned briefly in section 3.3 are explained further. They either required some manual work, or were evaluated in more detail for this bachelor thesis.

4.1.1 Temporal alignment of the EEG and MEG data

As briefly mentioned in section 3.3, the raw EEG and raw MEG data had to get temporally aligned. This was necessary, to eventually align the EEG data and the audio files. The MEG recorded the audio stimulus on a separate analogue channel, which enabled the precise alignment of the MEG data and the presented audio. The MEG and EEG data were recorded on two different computers and the measurements were started manually. Therefore, both measurements referenced separate clocks that exhibit slight differences. These differences cause the datasets to temporally diverge for longer measurements, like in this case. During the measurements, trigger pulses were sent which were captured by the EEG and MEG. With a software that was developed by a cooperation partner at the University Hospital Erlangen, the times, at which these trigger pulses occurred, can be found within each data set. Then the software uses those trigger pulses to compute the rates with which both signals diverge over time. The sampling rate of the EEG data is then adjusted to fit to the MEG dataset. This had to be done for each of the 50 subjects separately. Since the software does not work automated, the activation of the used functions, as well as the transfer of the found trigger times to

use for the adjustment of the EEG sampling rate, had to be done manually for each participant.

The temporally aligned EEG and MEG data were saved together in the original MEG dataset file. Therefore, the two datasets had to be separated again afterwards. To do so, a MatLab script was again provided by the research group. For each subject, the two datasets were detached and then saved separately in two files. Now, the digitized channel positions of the EEG electrodes on the head of the participant, that are used for the spatial alignment of the EEG data (see section 3.3), had to be manually extracted from the new EEG data file within the software. The channel positions were then saved in an Excel file for each subject individually. As some of the positions were sometimes wrongly formatted, each Excel table was then also examined again and respectively adjusted. Those last two steps, of manually extracting and separately saving the digitized channel positions for each participant are the reason, why the detaching process could not be done in an automated way either.

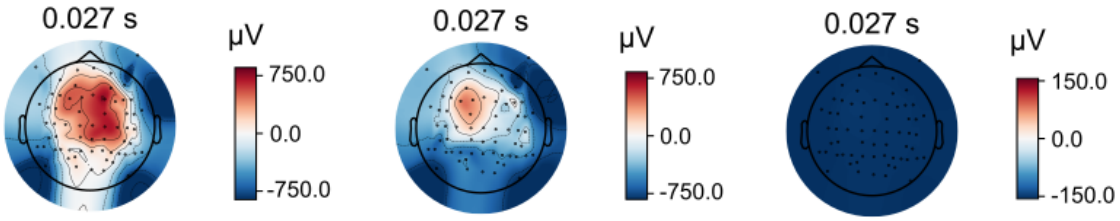
4.1.2 Selection of the referencing option

After the spatial alignment and frequency filtering mentioned in section 3.3, a reference potential had to be selected, since the EEG measures the potential differences between each electrode and a reference electrode. Generally, a useful reference signal does not capture the fluctuations in the electric potential that are caused by neural activity, but picks up those, which occur due to environmental noise, for instance. Thereby, as the potential value of the reference electrode is subtracted from the measurement electrodes, environmental disturbances in the signals are already removed from the data. Considering these requirements for the reference electrode, it seems reasonable that in practice the reference electrode is often placed close to the subject's head to measure the environmental disturbances here, but still far enough away from the source of the neural activity which should be measured. The placement of the ideal reference electrode is still subject of ongoing research [35].

In the measuring process of the data used here, no specific electrode was designated as the reference, which makes this step in the preprocessing necessary. The referencing was applied using the MNE-Python [14] 'set_eeg_reference()' function. One of the parameters of that function is the 'ref_channels' parameter, with which the chosen reference electrode can be set. In the selection process, the three most common referencing schemes were tried. In the first one, 'ref_channels' was set to 'average'. By doing so, a virtual EEG electrode is generated, with its electrical potential values resulting from averaging over the potentials of all 64 available real EEG channels. The second used option was 'ref_channels=[Cz]'. Here, the voltage values of the 'Cz' electrode, which is placed directly in the center of the scull (see Figure 2.4), are used as potential references for the other electrodes. The last scheme that was tried, was using the mean of multiple electrodes as a reference. In this case the mean of the mastoid electrodes 'M1' and 'M2' (see Figure 2.4) was used. To find out, which referencing method would be the most suitable for the EEG data in this work, the rest of the preprocessing had to be done, as well as the calculation and plotting of the TRFs and topoplots

(two dimensional topographic plot of the EEG-measured potential field). Ideally, this would have been done for all 50 subjects, for all three referencing schemes. However, after evaluating the three options for only a few subjects, the 'average' reference was chosen for the rest of the analysis.

The reasons therefore were the most sensible looking topoplots and TRFs with the 'average' reference, as well as the least unwanted alterations to the EEG data itself. This is illuminated further in the following, by the example of the data of subject 10. For the reference option of using the mean of the two mastoid electrodes 'M1' and 'M2', the plot of the EEG channels only showed completely disturbed signals. Therefore, the topoplot shown in Figure 4.1c, that resulted from the data with this referencing scheme, is practically meaningless and does not represent the measured potential field in any way. The plot of the EEG data of subject 10 resulting from the 'Cz' referencing option looked reasonable. However, the topoplot at the latency of 27 ms shown in Figure 4.1b is not as expressive as the one resulting from the 'average' referencing option shown in Figure 4.1a. The topoplot in Figure 4.1a depicts a potential field, that could be the result of cortical neural activity, which can not be stated for Figure 4.1b. Since the potential field in Figure 4.1b does not resemble a typical potential field arising from a certain neural activity, it is not as meaningful. Furthermore, the 'average' reference option has been worked with the most within the research group and thus far provided satisfactory results, which was another reason for choosing this method of referencing for the data in this work.



(a) 'average' (b) 'Cz' (c) Mean of 'M1' and 'M2'

Figure 4.1: Comparison of the topoplots that result from the three referencing schemes tested on the data of subject 10. (a) The 'average' reference option leads to the most sensible looking topoplot, as the shown potential field could arise from cortical neural activity. (b) The single electrode reference option with the 'Cz' electrode leads to a potential field that can not be assigned to a certain neural activity, making it not as meaningful. (c) Using the average of the mastoid electrodes 'M1' and 'M2' as reference potential leads to a practically meaningless topoplot, that does not represent the measured potential field.

4.1.3 Selection and handling of 'bad channels'

The next step was to plot the 64 EEG channels, to manually select those, which showed artifacts or no signal at all. Such artifacts can be recognized by either an overdrive or the loss of the signal that was captured by a certain electrode at some point of the measurement. They are caused, for instance, by the unwanted shift of the electrode on the head or a loss of electrical conductivity due to smudging of the contact gel beneath the electrode. Generally, artifacts are any variations in the EEG data that are not caused by neural activity. Artifacts that arose from eye movement or heart activity were specifically excluded later on in the preprocessing (see section 3.3), as they were captured by the electrooculography and electrocardiography. So, for each of the 50 participants, the measured EEG data of around $40\,\mathrm{min}$ of all channels was manually observed. In the next step, the 'bad channels' showing artifacts or no signal were marked, noted and interpolated (mne.io.Raw.interpolate_bads()) with respect to their neighboring 'good channels'. During the actual measurement of the data, the contact of some electrodes to the scull decreased, or was sometimes lost completely. These electrodes were noted for each participant and the notes were available to me. However, since they did not include all of the 'bad channels' that were found in the observation process, this manual selection was important to ensure cleaner data with as little distortion as possible. As an example, the plot of some of the EEG channels of subject 10 is depicted in Figure 4.2. The image shows the step of observing the data (at $\sim 582\,\mathrm{s}$ to $\sim 592\,\mathrm{s}$) and selecting the 'bad channels', which are here already marked (light grey coloring) to be interpolated afterwards. The channels 'FT7' and 'TP10' show clear irregularities compared to the signals of the other electrodes.

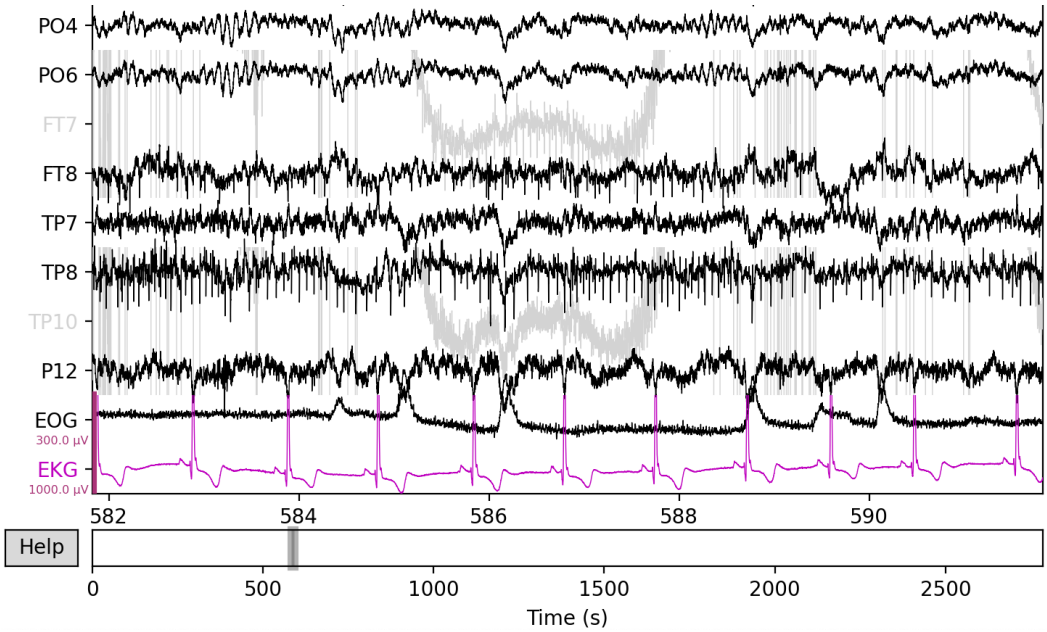
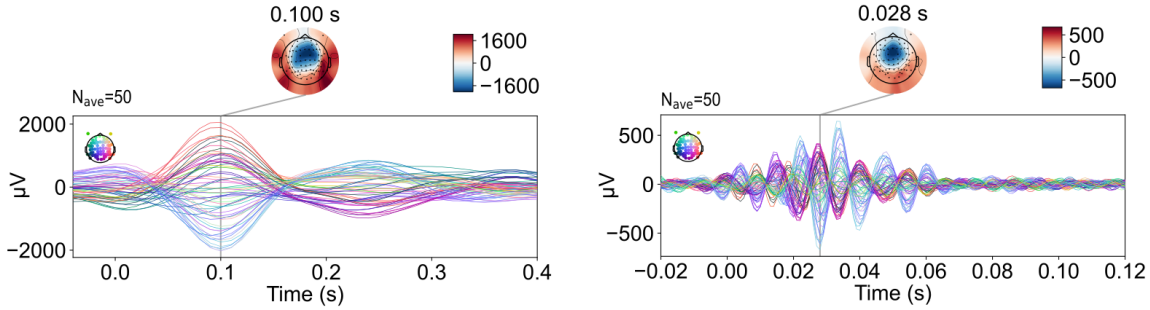


Figure 4.2: Selection of 'bad channels' for the data of subject 10. The signals of the electrodes 'FT7' and 'TP10' show distinct irregularities compared to their neighbouring channels. Therefore, they were marked (light grey coloring) to be interpolated afterwards.

4.2 Other acoustic features

As mentioned in section 3.4, the fundamental waveform and the envelope modulation of the higher harmonics of the fundamental frequency were chosen as the acoustic features of the audiobooks narrated by the lower pitch speaker to use for the analysis of the EEG data. However, before settling on these, other acoustic features were considered as well. The motivation for their consideration, the process of their extraction, as well as a discussion about why they were discarded in the end is given in the following.



(a) (b)

Figure 4.3: Population-averaged TRFs of the attended mode of all 50 participants with the (a) low-frequency envelope and the (b) high-frequency envelope used as acoustic features, respectively. (a) The envelope of the audiobooks, as well as the EEG data, were filtered between 1 Hz and 20 Hz. (b) The envelope of the audiobooks, as well as the EEG data, were filtered between 60 Hz and 120 Hz.

4.2.1 Low-frequency envelope

In addition to the brain responses to high-frequency acoustic features, such as the fundamental waveform and the envelope modulation, research has shown, that low-frequency acoustic features also induce neural responses. For instance, Giovanni M. Di Liberto et al. [9] used the low-frequency amplitude envelope (1 Hz-15 Hz) of the speech signal to show that the measured EEG data reflects the processing of phonemes within continuous speech. Phonemes are sets of speech sounds (phones) in a language, that would, if exchanged for another phoneme in a word, alter the meaning of that word. In their study, they also investigated the EEG data specifically within the delta (1-4 Hz) and theta (4-8 Hz) bands, which also reflected cortical brain activity. The low-frequency neural responses are also currently worked on by other members within the research group with the MEG data that was recorded together with the here used EEG data. For these reasons, this low-frequency analysis was considered and tried for the EEG data as well.

At first, the acoustic envelope of the presented audiobooks of the lower pitch speaker needed to be computed. Both, the audiobook that was attended and the one that was ignored, were provided to me

as five separate files. For each of these five files, the envelope was extracted by taking the absolute values of the Hilbert transform of the audio signal. Then, a Butterworth bandpass-filter was applied to the envelope. The cutoff frequencies were chosen to be $1\,\mathrm{Hz}$ and $20\,\mathrm{Hz}$ for the evaluation within the low-frequency broadband, $1\,\mathrm{Hz}$ and $4\,\mathrm{Hz}$ for the evaluation within the delta band and $4\,\mathrm{Hz}$ and $8\,\mathrm{Hz}$ for the evaluation within the theta band, respectively. Then, the filtered envelopes were downsampled from $1000\,\mathrm{Hz}$ to $100\,\mathrm{Hz}$, to reduce the computation time in the further processing of the data. Lastly, the five separate envelopes were joint back together. This process resulted in two envelopes for each filtered frequency band, one for the attended mode and one for the ignored mode of the lower pitch speaker, respectively.

To now calculate the TRFs with the differently filtered envelopes as acoustic features, the EEG data was also downsampled to $100\,\mathrm{Hz}$ and bandpass-filtered within the respective frequency ranges of $1\,\mathrm{Hz}$-$20\,\mathrm{Hz}$ for the low-frequency broadband, $1\,\mathrm{Hz}$-$4\,\mathrm{Hz}$ for the delta band and $4\,\mathrm{Hz}$-$8\,\mathrm{Hz}$ for the theta band. Here, only one acoustic feature was used for the computation of the TRFs, respectively, other than the two acoustic features used in Equation 3.1. Due to their low-frequency nature, using the envelopes as acoustic features also leads to later neural responses than the high-frequency features. This phenomenon is called 'neural speech tracking'. It describes, how neural responses in the brain are time-locked to the acoustic features of a continuous speech stimulus [4]. It is thought, that the brain matches the rhythm of the external input in its processing. Therefore, a slower acoustic feature of the speech stimulus results in a later brain response. High-frequency acoustic features, such as the fundamental frequency and the envelope modulation, thereby lead to 'faster' brain responses. Furthermore, the neural responses, that can be analysed by using the low-frequency envelope as an acoustic feature for the TRF computation, originate in the cortex of the brain [4]. As an example, Figure 4.3a shows the population-averaged TRF over all 50 participants in the attended mode, where the low-frequency envelope was used as an acoustic feature. In this case, the audio envelope and the EEG data were filtered between $1\,\mathrm{Hz}$ and $20\,\mathrm{Hz}$. The topoplot suggests a cortical response at $100\,\mathrm{ms}$.

While this method of analysing the EEG data brought satisfying and interesting results, we lastly decided to rather proceed with the analysis using the fundamental waveform and the envelope modulation of the lower pitch speaker as acoustic features. Since the corresponding MEG data had already been analysed with those two features, it seemed more sensible for the research group to also analyse the EEG data with these first, to compare the results obtained by the two different measuring techniques.

4.2.2 High-frequency envelope

Additional to the brain responses to the low-frequency envelope of the audiobooks, the audio envelope within a higher frequency range was considered as an acoustic feature for the analysis of the EEG data as well. This idea arose due to very recent findings within the research group. For the

'Auditory EEG Decoding Signal Processing Grand Challenge' of the ICASSP (International Conference on Acoustics, Speech, and Signal Processing) 2024, Thornton et al. [33] used high-frequency gamma-band responses to the audio envelope to match a short segment of EEG data to the right one of several given speech segments.

To extract the envelope of the audiobooks of the lower pitch speaker, a similar procedure was implemented as described above. For each of the five audio files, for the attended and the ignored mode, respectively, the envelope was computed by taking the absolute values of the Hilbert transform of the audio signal. Then, again, a Butterworth bandpass-filter was applied, however, this time with a lower-cutoff frequency of $60\,\mathrm{Hz}$ and a higher-cutoff frequency of $120\,\mathrm{Hz}$. These boundaries were chosen, as they reflect the frequency range of the fundamental frequency of the lower pitch narrator. The lower edge was chosen to be $60\,\mathrm{Hz}$ instead of the $70\,\mathrm{Hz}$ used for the fundamental waveform and the envelope modulation, as it has been shown within the research group, that the quality of the results improves by choosing a slightly broader frequency range when working with this still new acoustic feature. Contrary to the low-frequency envelopes, the high-frequency envelopes did not have to be downsampled, but remained at a sampling frequency of $1000\,\mathrm{Hz}$. As the neural responses occur earlier than for the low-frequency envelope (see next paragraph), the observed latency range did not have to be as broad, making it possible to use a higher sampling rate without a large amount of computation time. For both attentional modes, the five respective envelopes were joint back together, resulting in two audio envelopes of the lower pitch speaker.

For the TRF computation, the EEG data now also did not have to be downsampled, but retained its sampling frequency of $1000\,\mathrm{Hz}$. Again, only one acoustic feature was used for the TRF calculation, contrary to Equation 3.1. Due to the high-frequencies of the audio envelope, the responses in the brain occur at smaller latencies than for the low-frequency envelope. The reason therefor is again the above (subsection 4.2.1) described 'neural speech tracking'. Again, as an example, Figure 4.3b shows the population-averaged TRF over all 50 participants in the attended mode, where the high-frequency envelope was used as the acoustic feature. The audio envelope and the EEG data were filtered between $60\,\mathrm{Hz}$ and $120\,\mathrm{Hz}$. The topoplot possibly suggests a cortical response at $28\,\mathrm{ms}$.

Although this was an interesting test of the only recently, within the research group, investigated option of the analysis of EEG data, it did not lead to the results hoped for. Together with the same reason of comparability as mentioned above, the high-frequency envelope was discarded as an acoustic feature for the further analysis of the EEG data in this bachelor thesis.

4.3 Evaluation of the regularization parameter λ

As mentioned in section 3.5, five-fold cross validation was used during the process of the subject-wise computation of the TRF coefficients, to choose the most adequate regularization parameter λ for all of the 50 participants. The cross validation was performed between the acoustic stimuli (fun-

damental waveform and envelope modulation) and the measured EEG data, to investigate which λ provided the most accurate prediction of the EEG data capturing the brain response to the acoustic stimuli. The model therefore needed suggestions for the possible regularization parameters, which were chosen to be 12 logarithmically organized values, ranging from 0 to 10000. While the Python code for the cross validation was provided to me through the script for the TRF computation, the outcome of the cross validation process still had to be evaluated.

Said outcome were 12 evaluation scores for each participant, that depicted how well the linear forward model, used for the TRF estimation, could predict the EEG data from the acoustic stimuli, depending on the chosen regularization parameters. The scores were then plotted over the suggested regularization parameters for each of the subjects. Additionally, the maximum score and its corresponding λ were identified, again for all 50 participants. As the maximum score reflects the best model capability of the respective λ, this provided an optimal regularization parameter for each subject. Moreover, the subject-wise plots were checked for peaks with a negative score. A negative evaluation score would have been a reason to exclude the data of a participant from the further analysis. This would have implied, that the regularization and the linear forward model generally do not work for the data of the participant. In the end, the data of all 50 subjects could be used for the further analysis, as no negative evaluation scores resulted from the cross validation.

Afterwards, by counting the number of times each λ was the optimal regularization parameter for a participant, it was possible to determine which λ would work best for which number of subjects. The 12 λ were each assigned to an index from $[0, 1, 2, ...11]$. Then, each index was multiplied by the number of times the respective lambda was deemed optimal. These products were added up and the resulting sum was divided by the total number of subjects. In doing so, a mean index was found, which correlated to the regularization parameter of $\lambda = 5$. This was then chosen as the regularization parameter to use for each subject for the further analysis, as it would give on average the most satisfying results for all participants.

In addition to the above described evaluation, it was tried to work with individual regularization parameters. So, for every subject, the optimal λ was noted and used for the plotting of the TRFs. While this worked well for the single-subject TRFs, it provided an issue when computing the population-averaged TRF. Since the data of each subject was now regularized to a different extent, the single-subject TRF coefficients needed to be zscored (scipy.stats.zscore()) first, to receive data of the same scale. This altered the resulting population-averaged TRF. It furthermore did not seem to provide far better results than the use of one collective λ for the data of all participants. While testing this method, I unfortunately could not find any research that attempted to do the same and that could thus have provided some insightful information. In the end, this way of regularizing the TRF coefficients therefore was discarded.

4.4 Alignment check

To possibly improve the results of the analysis of the EEG-measured neural responses of musicians and non-musicians in the competing speaker scenario, an 'alignment check' was implemented. With this procedure, it was attempted to create a more precise temporal alignment of the captured EEG data of each participant and the acoustic features of the presented auditory stimulus. This would subsequently lead to a more accurate representation of the temporal delay and the strength of the brain responses when observing the population-averaged TRF. To find out, whether the EEG data and the acoustic features were perfectly aligned, it needed to be tested, whether the EEG and MEG data were perfectly aligned (see subsection 4.1.1). The following process was performed for all of the 50 subjects.

First, an artificial EEG stimulus channel was created. Therefore, the times of the trigger pulses were selected with the software, that was also used for the alignment process of the EEG and MEG data in subsection 4.1.1. Now, one of the 64 EEG channels was overwritten with data that generally showed no signal, but peaks at the specific trigger times. This step was necessary, since there was no designated EEG channel that only captured the trigger pulses in the first place. This dataset with the new EEG channel was saved as a separate file. Now, again with the software used for the alignment process in subsection 4.1.1, this new EEG file was temporally aligned with the raw MEG data. Afterwards, the datasets were separated again with the provided MatLab script. This resulted in one file containing the MEG data, where the trigger pulses were originally captured by a designated channel, and one file containing the EEG data, that was now temporally aligned to the MEG data and that now also included one channel only presenting the trigger pulses.

The stimulus channels of both separate files were then plotted together into one figure over the sample points. It was now possible to determine the shift between the trigger pulse peaks of the MEG and EEG channel in samples and thereby evaluate the accuracy of the MEG/EEG alignment software.

Ideally, there of course would have been no shifts at all, proving a perfect alignment obtained with the software that was therefor used. However, for almost all subjects, a slight temporal shift was found between the EEG and MEG data. While the peaks in the EEG data were always located a few sample points later than those in the MEG data, the number of samples between them was not the same for all participants. Therefore, the shift in samples was noted for each subject and individually considered in the later analysis. These shifts ranged from 1 to 5 sample points.

4.5 Magnitude computation

For a first representation of the timing and strength of neural responses to an auditory stimulus, population-averaged TRF plots are a useful tool. However, for a more detailed investigation of the

data, especially for the comparison of the impact of different criteria, such as the musicality of the participants in this case, or for a statistical analysis, the TRFs need to be processed further. Therefore, the magnitudes of the TRFs were computed. During the work on the data used for this bachelor thesis, an interesting difference was found in the magnitudes, depending on the order of the averaging steps in the magnitude calculation process. The two different approaches are explained below, while the different results they provide are depicted in section 5.4.

The version of magnitude computation that was tried first, is the common process used within the research group. For each subject, the absolute values of each EEG channel of the TRF are taken. Then, still for each subject separately, the channel average over all of the 64 EEG channels is computed. Lastly, to enable the analysis on the population level, the population average is computed over all 50 single-subject magnitudes.

By working in this above described order, the resulting magnitude plot does not resemble the general shape of the population-averaged TRF. Since some interesting differences between the musicians and non-musicians, that seemed to be present in the TRFs, were lost with this process, another way of calculating the magnitude was tried. Now, the population average was computed over the single-subject TRFs first. Then, the absolute values of each EEG channel were taken, followed by the computation of the channel average. In doing so, the magnitude plot actually resembled the population-averaged TRF. For this reason, this second approach to the magnitude computation was chosen. A comparison of the magnitude plots resulting from both orders of computation is given in section 5.4. The possible reasons behind the differences of the two approaches are discussed in section 5.4 and section 6.1.

4.6 Statistical analysis

For a quantitative evaluation of the TRFs, statistical tests were carried out to analyse the significance of differences between the attended and ignored modes, as well as between the responses of the musicians and non-musicians.

For both relations, t-tests were employed in order to assess the thereby obtained p-values. The t-test is a test for the null hypothesis suggesting that two compared samples, or sets of samples, have the same expectation values, or average values, and therefore are drawn from the same distribution [10]. The p-value gives the probability for obtaining the same, or more extreme, t-test statistics under the premise, that the null hypothesis is true. If the p-value is lower than a certain percentage, the null hypothesis can be dismissed and the difference between the compared samples can be deemed significant. If the p-value is larger than that percentage, the results could have also occurred by chance and the null hypothesis therefore can not be rejected. For this analysis, the threshold was chosen to be 5%, as it is commonly used in this field of research, as well as within the work group. The t-tests were employed by using the 'scipy.stats' module in Python.

Three latencies were chosen, at which the t-tests were performed. The aim was to choose one latency, where a subcortical response is expected ($\sim 5\,\mathrm{ms}$ - $15\,\mathrm{ms}$) and one, where a cortical response is expected ($\sim 15\,\mathrm{ms}$ - $35\,\mathrm{ms}$). Since there also seemed to be a neural response even later, a third latency was chosen, as well. To pick the three latencies, the population-averaged TRF when the lower pitch speaker was attended was observed, together with the topoplots at certain latencies, as well as the corresponding population-averaged magnitude plot, which was computed via the second approach described in section 4.5. With the topoplots, it could be suggested, in which region of the brain the response occurred. With the TRF and magnitude plot, latencies could be picked, at which a peak was visible in general. This needed to be done for the fundamental waveform and the envelope modulation. It resulted in the latencies of $8\,\mathrm{ms}$, $27\,\mathrm{ms}$ and $53\,\mathrm{ms}$ for the fundamental waveform and $5\,\mathrm{ms}$, $23\,\mathrm{ms}$ and $47\,\mathrm{ms}$ for the envelope modulation. Again, judging by the corresponding topoplots, at the smallest latencies, the responses are assumed to originate in the subcortex and at $27\,\mathrm{ms}$ or $23\,\mathrm{ms}$, the responses should occur in the cortex. The $53\,\mathrm{ms}$ or $47\,\mathrm{ms}$ was additionally chosen due to interesting observations in the TRF.

While the population-averaged TRF and magnitude were used for the latency-selection, for the t-tests, the single-subject magnitudes were necessary. So for each subject, the absolute values of the single-subject TRF were taken, followed by the average over all 64 EEG channels. Now, the magnitude values at the respective three latencies were selected. With these single-subject magnitude values at the certain latencies, the t-tests were performed. This was done for both acoustic features and both attentional modes.

4.6.1 Significance of attended versus ignored

For the comparison of the responses in the attended and ignored mode of the lower pitch speaker, the t-test was performed with related samples (scipy.stats.ttest_rel()), since there was data of both modes for each subject. At each of the three latencies, the magnitude values of each subject in the attended mode were compared to the respective magnitude values of the same subject in the ignored mode. This was performed for both acoustic features.

4.6.2 Significance of musicians versus non-musicians

For comparing the responses of the musicians to those of the non-musicians, the t-test was performed with independent samples (scipy.stats.ttest_ind()), since each subject could be assigned to only one of the two groups. At each of the three latencies, the single-subject magnitude values of the musicians in the attended mode were compared to the single-subject magnitude values at the respective times of the non-musicians in the attended mode. This was again performed for the fundamental waveform and the envelope modulation, as well as for the ignored mode, respectively.

Additionally, the responses in the attended and ignored mode were compared within the groups of the musicians and the non-musicians. So, at the three chosen latencies, the single-subject magnitude values of the musicians in the attended mode were compared to the single-subject magnitude values of the musicians in the ignored mode. This was done for the non-musicians, respectively, as well as for both acoustic features.

5 Results

In this chapter, the results found by the analysis of the EEG-measured neural responses to the lower pitch narrator in the competing-speaker scenario are presented. In the first section, the attentional modulation of the brain response is investigated on the population level, as well as on a single-subject level. The second section covers the influence of musical training on the brain responses, by comparing the magnitudes of the TRFs of the musicians to those of the non-musicians. Afterwards, the impact of attention to the narrator and musicality of the participants are evaluated together, by comparing the magnitudes of the TRFs of the attended and ignored mode within the groups of musicians and non-musicians. Lastly, the impact of the order of the steps in the calculation of the magnitudes of the TRFs (see section 4.5) is shown.

While, out of time reasons within this bachelor thesis, no statistical tests on whether the neural responses are significant compared to a noise level were carried out, a brain response can be recognised by a clear peak in the presented TRF and magnitude plots.

5.1 Attentional modulation of the brain response

This section presents the results found by analysing the attentional modulation of the brain responses to the lower pitch speaker. At first, the results on a population level are shown. To get a more detailed understanding of these results, the data is then investigated on a single-subject level.

5.1.1 On the population level

To analyse the data on a population level, the single-subject TRFs were computed and then averaged over all 50 participants. To investigate the influence of attention on the brain responses, this was carried out for the data, when the lower pitch speaker was attended and when he was ignored. Afterwards, the magnitudes of the population-averaged TRFs were determined, according to the second approach in section 4.5. Now, the magnitudes on the population level in the attended mode can be compared directly to those in the ignored mode. Additionally, statistical tests were employed to examine the significance of the differences between the attended and ignored mode of the lower pitch speaker. All of the above mentioned steps were employed to the TRFs of both acoustic features separately. Note that for the depicted magnitude plots, the population average was calculated first, while the statistical analysis uses single-subject magnitude values. This leads to a possible visual mismatch of the obtained p-values and the respective plots. The issue is explained further in section 5.4 and

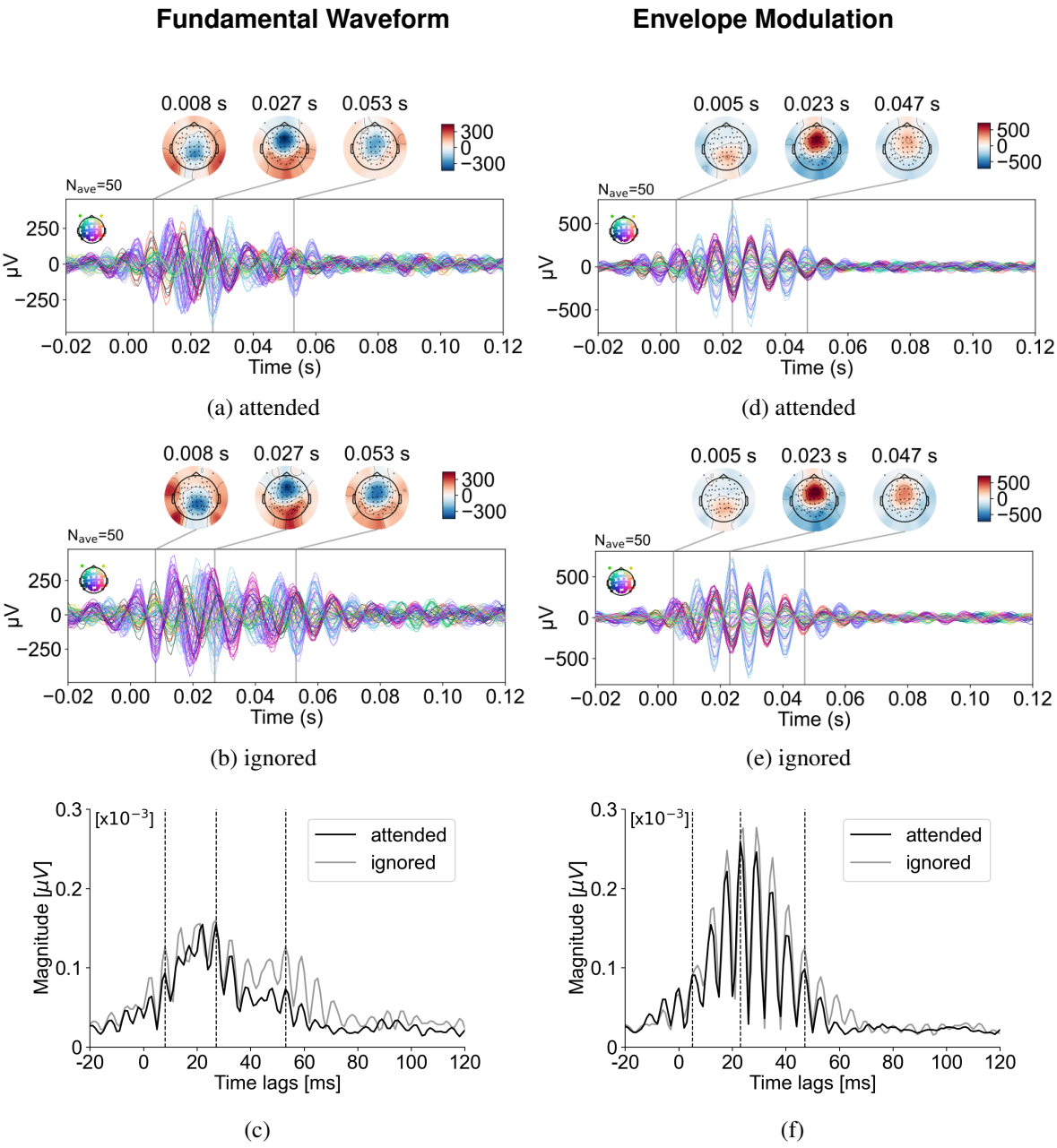
subsection 6.1.4.

Figure 5.1a and Figure 5.1b show the population-averaged TRFs of the fundamental waveform in the attended mode and in the ignored mode of the lower pitch narrator, respectively. The 64 EEG channels are marked in different colors and can be assigned to the corresponding electrodes and their specific placements on the head via the legend in the upper left corner of the TRF plot. The depicted topoplots model the potential field on the two-dimensional scull at the given latencies. The colorbar on the right side, next to the topoplots, shows the potential values and their corresponding color coding in the topoplots. The specific latencies were chosen according to section 4.6 and show the times at which the t-tests were performed.

The TRF in Figure 5.1a shows a neural response in the latency interval between approximately $0\,\mathrm{ms}$ and $60\,\mathrm{ms}$, with the greatest amplitudes between approximately $15\,\mathrm{ms}$ and $30\,\mathrm{ms}$. This also aligns with the topoplot at $27\,\mathrm{ms}$ showing a stronger coloring and therefore higher potential differences than the topoplots at $8\,\mathrm{ms}$ ans $53\,\mathrm{ms}$. In Figure 5.1b, the TRF shows a brain response between the latencies of approximately $0\,\mathrm{ms}$ and $70\,\mathrm{ms}$. The highest peaks approximately lie between $10\,\mathrm{ms}$ and $35\,\mathrm{ms}$. Therefore, the entire interval of neural response, as well as the region of strongest response, is broader for the ignored than for the attended mode. Again, the topoplot at $27\,\mathrm{ms}$ shows the strongest coloring of the three depicted latencies. However, both topoplots at $8\,\mathrm{ms}$ and $53\,\mathrm{ms}$ show stronger potential differences in the ignored than in the attended mode.

Figure 5.1d and Figure 5.1e show the population-averaged TRFs of the envelope modulation in the attended and ignored mode of the lower pitch speaker, again with topoplots depicting the potential field on the head at the latencies of the conducted t-tests. The envelope modulation seems to generally cause a stronger brain response than the fundamental waveform, which can be seen by comparing the y-axis of the TRF plots and the colorbar of the topoplots of the two acoustic features.

The TRF in Figure 5.1d shows a neural response approximately between $0\,\mathrm{ms}$ and $60\,\mathrm{ms}$, aligning with the interval of the attended mode of the fundamental waveform. For the envelope modulation in the attended mode however, the highest peaks appear to be between $10\,\mathrm{ms}$ and $40\,\mathrm{ms}$. The topoplot at $23\,\mathrm{ms}$ shows a stronger coloring, correlating with greater potential values, than the topoplots at $5\,\mathrm{ms}$ and $47\,\mathrm{ms}$. The interval of brain responses in the TRF plot in Figure 5.1e can be approximated to stretch between $0\,\mathrm{ms}$ and $60\,\mathrm{ms}$. The region of greatest amplitudes equals the one for the attended mode of the envelope modulation, reaching from approximately $10\,\mathrm{ms}$ to $40\,\mathrm{ms}$. So, the total latency interval of neural responses is approximately the same in the ignored and in the attended mode of the envelope modulation, as well as the region of strongest brain responses. Again, the topoplot at $23\,\mathrm{ms}$ shows the strongest potential differences out of the three latencies. Compared to the topoplots in Figure 5.1d, the ones at $5\,\mathrm{ms}$ and $23\,\mathrm{ms}$ seem similar, while the one at $47\,\mathrm{ms}$ possibly shows a slightly stronger coloring.



(a) attended

(d) attended

(b) ignored

(e) ignored

(c)

(f)

Figure 5.1: Attentional modulation of the brain response to the fundamental waveform (a-c) and the envelope modulation (d-f) on the population level. (a) Averaged TRF of all 50 subjects in the attended mode. (b) Averaged TRF of all 50 subjects in the ignored mode. (c) Magnitude comparison of the attended and ignored mode. The dotted lines mark the 8 ms, 27 ms and 53 ms where the t-tests were performed. At all three considered latencies, the differences can be deemed significant. (d) Averaged TRF of all 50 subjects in the attended mode. (e) Averaged TRF of all 50 subjects in the ignored mode. (f) Magnitude comparison of the attended and ignored mode. The dashed vertical lines mark the 5 ms, 23 ms and 47 ms where the t-tests were performed. The differences are significant at 23 ms and 47 ms.

In all four of the above described plots, a spatial shift in the strongest regions of the potential fields can be observed over different latencies in the respective topoplots. While the most active area of the brain (blue spot) at $8\,\mathrm{ms}$ in the topoplots in Figure 5.1a and Figure 5.1b can be located more towards the back of the head, it shifts towards the front of the head at $27\,\mathrm{ms}$. This also applies for the most active area (red spot) at $5\,\mathrm{ms}$ and $23\,\mathrm{ms}$ in Figure 5.1d and Figure 5.1e. Here however, the most active brain area at $5\,\mathrm{ms}$ can only be assumed, since the neural response does not seem to be as strong in general at this early latency for the envelope modulation. Note that the reverse coloring of the potential fields with the fundamental waveform compared to the envelope modulation is due to different polarity and has no meaning for the general interpretation of the results.

These spatial shifts in the topoplots suggest responses of different brain areas at the first and second chosen latency. The topoplots at $8\,\mathrm{ms}$ and $5\,\mathrm{ms}$ hint to an activation in the subcortical part of the brain. Neural activity can be located in the areas of the mastoid EEG channels 'M1' and 'M2' (darker red spots behind the ears), which is a common pattern of subcortical brain responses. At $27\,\mathrm{ms}$ and $23\,\mathrm{ms}$, the topoplots allow for the assumption of cortical responses, due to the strongest activity being located at the front half of the head.

The topoplots at $53\,\mathrm{ms}$ and $47\,\mathrm{ms}$ show a slight shift of the most active area to the back of the head again, compared to the topoplots at $27\,\mathrm{ms}$ and $23\,\mathrm{ms}$, respectively. As they lie at the center of the head, it is rather difficult to assign the neural responses to a specific area of the brain.

For a direct comparison, the plots in Figure 5.1c and Figure 5.1f show the magnitudes of the TRFs of the attended and ignored modes of the fundamental waveform and the envelope modulation, respectively. The computation of the magnitudes followed the description of the second approach in section 4.5. The absolute values of each EEG channel of the population-averaged TRFs pictured above the magnitude plots in Figure 5.1 were taken, followed by the channel average over all 64 channels. In both plots, the dashed vertical lines mark the latencies at which the t-tests were performed.

Figure 5.1c shows overall similar magnitudes for the attended and ignored mode of the fundamental waveform. Especially until about $35\,\mathrm{ms}$, the magnitude curves mostly overlap, with some peaks of the ignored response being greater than the attended one. Between $35\,\mathrm{ms}$ and $80\,\mathrm{ms}$ however, the response in the ignored mode shows higher magnitudes than the response in the attended mode. To test the significance of these differences, t-test were carried out with the respective single-subject magnitude values according to section 4.6. At $8\,\mathrm{ms}$, as well as at $27\,\mathrm{ms}$, the thereby obtained p-values are $p = 0.01$. At $53\,\mathrm{ms}$, the p-value was found to be $p < 0.001$. With all of these being smaller than the here chosen threshold of 5%, the differences of the magnitudes of the neural responses in the attended and ignored mode of the lower pitch speaker can be deemed significant at all three latencies

for the fundamental waveform. The p-values appear rather extreme compared to Figure 5.1c. The reason behind this discrepancy is illuminated in subsection 6.1.4.

In Figure 5.1f, the magnitudes of the TRFs of the attended and ignored mode of the envelope modulation show a similar shape for the entire latency interval. At the peaks between approximately $5\,\mathrm{ms}$ to $70\,\mathrm{ms}$, the brain response in the ignored mode seems to be slightly higher than in the attended mode. At the specific latencies, the t-tests resulted in p-values of $p = 0.08$ at $5\,\mathrm{ms}$, $p = 0.03$ at $23\,\mathrm{ms}$ and $p < 0.001$ at $47\,\mathrm{ms}$. Therefore, the differences between the magnitudes of the TRFs in the attended and ignored mode of the envelope modulation can be deemed significant at $23\,\mathrm{ms}$ and $47\,\mathrm{ms}$, since the p-values at these latencies are ≤ 0.05, and insignificant at $5\,\mathrm{ms}$, since the p-value is slightly above the threshold.

5.1.2 On a single-subject level

To further explore and understand the findings described above, the data was subsequently analysed on a single-subject level. Therefore, three subjects were picked, with each of them representing either 'bad', 'unexpected' or 'expected' data. What these adjectives mean is explained regarding the respective plots in the following and will thus be clear at the end of this section. Figure 5.2 depicts the magnitude plots of the TRFs of both acoustic features, comparing the attended (a) and ignored (i) mode for subject 13, subject 2 and subject 52, respectively.

As shown in Figure 5.2 **A** and **D**, no sensible TRFs and corresponding magnitudes could be derived from the measured EEG data of subject 13. For both attentional modes and both acoustic features, the magnitudes remain noisy for the entire latency interval and show no obvious peaks. This implies, that there was no measurable neural response in neither the attended, nor the ignored mode to any of the two acoustic features. Therefore, subject 13 is an example for a participant with results that are not satisfactory, or 'bad'. Such data was still considered for the analysis on the population level, since it is part of a realistic depiction of EEG measurement results.

Subject 2 would be an example for a participant with results that are 'unexpected'. In Figure 5.2 **B** and **E**, clear peaks can be observed for the attended and ignored mode and for both, the fundamental waveform and the envelope modulation. Moreover, the peaks are approximately in the same latency range as the peaks in the magnitude plots of the TRFs on the population level. However, the neural response in the ignored mode is clearly stronger than the neural response in the attended mode of the lower pitch speaker, especially for the fundamental waveform. This is very unusual and questionable in comparison to other studies on this subject and the current general understanding of attentional modulation of brain responses. A further discussion of this finding is provided in section 6.1. The data was again still considered in the population averages.

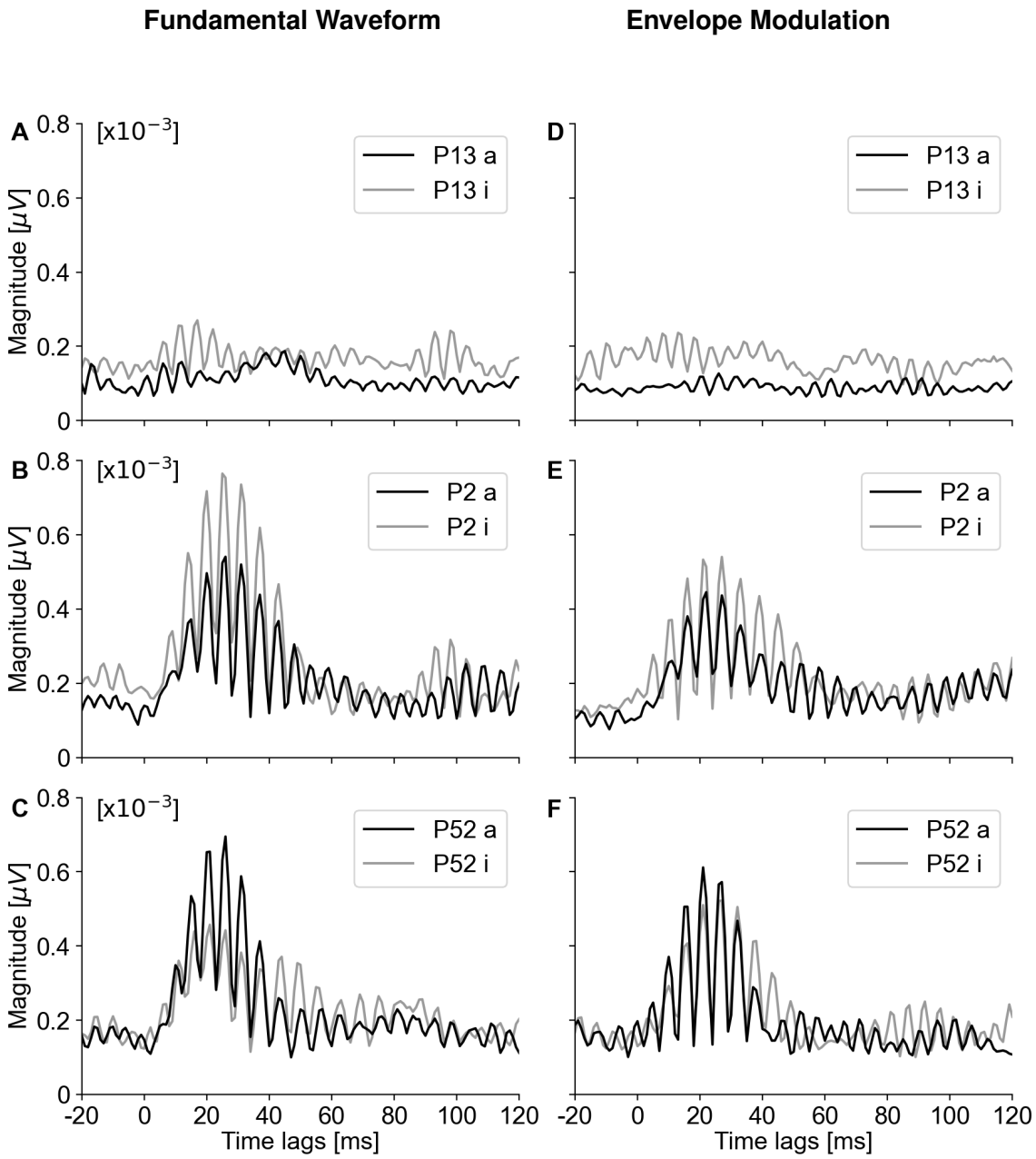


Figure 5.2: Attentional modulation of the brain response on a single-subject level. Magnitude comparison of the attended (a) and ignored (i) mode of **A** subject 13 - fundamental waveform, **B** subject 2 - fundamental waveform, **C** subject 52 - fundamental waveform, **D** subject 13 - envelope modulation, **E** subject 2 - envelope modulation, **F** subject 52 - envelope modulation. Subject 13 represents 'bad' results. Subject 2 represents 'unexpected' results, while subject 52 represents 'expected' results.

An example for 'expected' results is given by the data of subject 52, as shown in Figure 5.2 **C** and **F**. First of all, for both acoustic features, there are clear brain response peaks in the magnitudes, which are also approximately in the same latency range as the ones on the population level. Furthermore, for the most part, attending the lower pitch speaker resulted in a stronger neural response than ignoring him. Only some of the peaks of the ignored magnitude are higher than those of the attended one, between approximately 40 ms and 100 ms. The current consent on the question of attentional modulation of brain responses is, that there are usually stronger responses in the attended than in the ignored mode, which is why this subject represents the more expected result. This will again be discussed further in section 6.1.

5.2 Influence of musical training on the brain response

In this section, the influence of musical training on the brain responses is investigated. Therefore, the population-averages of the single-subject TRFs were taken within the groups of musicians and non-musicians. Afterwards, the absolute values of each EEG channel were calculated, followed by the channel average over all 64 of these, according to the second approach to the magnitude computation in section 4.5. Then, the magnitudes of the non-musicians were compared to those of the musicians, separately for both acoustic features and both attentional modes. This resulted in four comparison plots, which are depicted in Figure 5.3. The dashed lines mark the latencies, at which the t-test were conducted according to section 4.6. Note that for the t-tests the single-subject magnitude values were used and not the averaged magnitudes depicted in the plots. Therefore, the p-values may not match the magnitude plots visually. This issue of averaging in the magnitude computation process is discussed further in section 5.4 and subsection 6.1.4.

5.2.1 Influence of musical training on the timing of the brain response

Generally, the latency intervals in Figure 5.3, at which any neural responses can be seen, are similar to those of the population-averaged TRFs and magnitude plots in Figure 5.1. For the fundamental waveform (**A**,**B**), the brain responses above the noise level occur between approximately 0 ms and 60 ms for the attended mode and 0 ms and 70 ms for the ignored mode of the lower pitch speaker for the musicians and non-musicians, respectively. For the envelope modulation (**C**, **D**), the brain responses also lie between approximately 0 ms and 60 ms for the attended mode and 0 ms and 70 ms for the ignored mode for either one of the two groups.

By specifically observing the plots of the fundamental waveform in the attended and the ignored mode in Figure 5.3 **A** and **B**, the interval of the brain responses of the musicians still appears to be broader than the one of the non-musicians. The peaks in the magnitude curves of the musicians start to rise earlier than those of the non-musicians. Therefore, the musicians seem to have an earlier

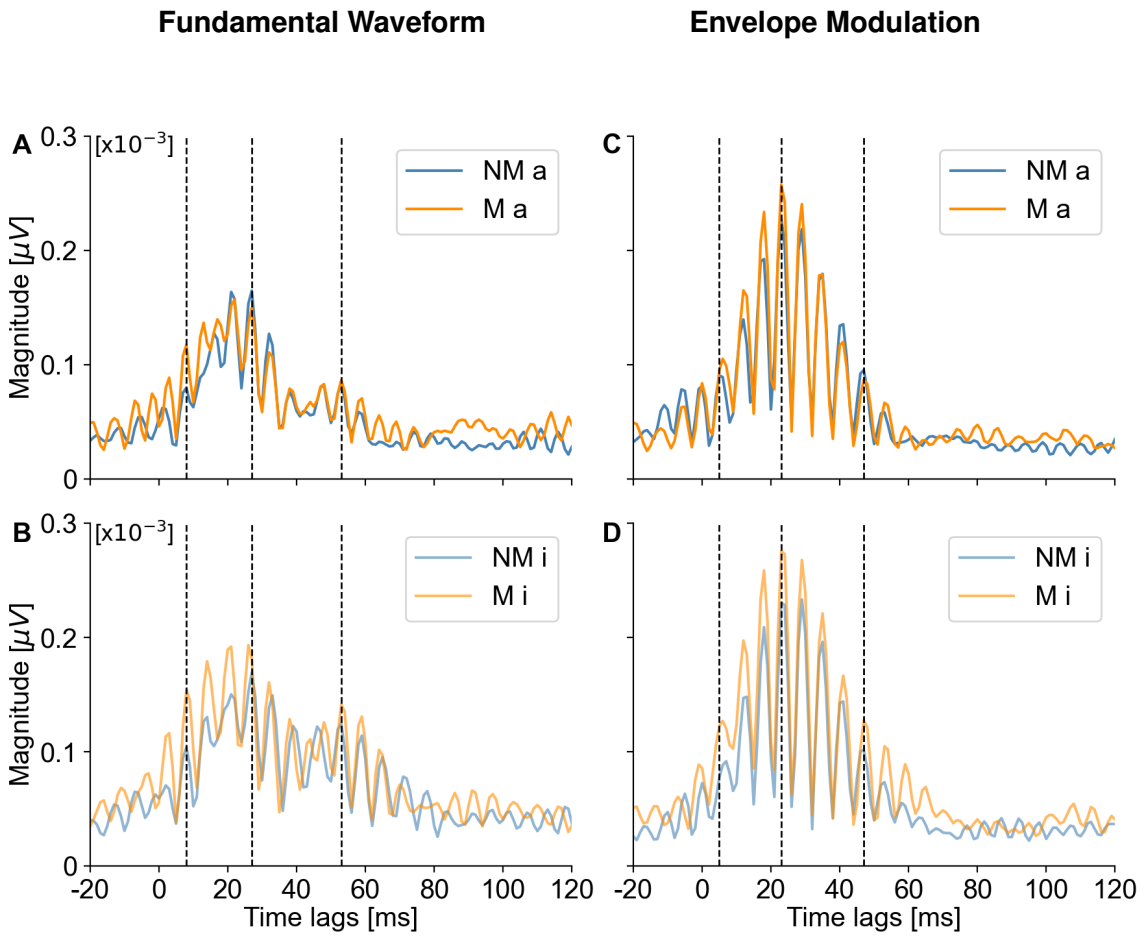


Figure 5.3: Influence of musical training on the brain response. Magnitude comparison of the TRFs of the musicians versus the non-musicians in the **A** attended (a) mode - fundamental waveform, **B** ignored (i) mode - fundamental waveform, **C** attended (a) mode - envelope modulation, **D** ignored (i) mode - envelope modulation. The dashed vertical lines mark the latencies at which the t-tests were performed. No significant differences were found for any attentional mode and any acoustic feature at the considered latency times.

neural response to the fundamental waveform than the non-musicians at those smaller latencies. This effect can not be observed for attended mode of the envelope modulation in Figure 5.3 **C**. For the ignored mode of the envelope modulation in Figure 5.3 **D**, the magnitude curve of the musicians seems to rise earlier than those of the non-musicians as well.

5.2.2 Influence of musical training on the strength of the brain response

In Figure 5.3 **A**, the magnitudes of the TRFs of the fundamental waveform in the attended mode of the musicians and the non-musicians mostly overlap. As mentioned above however, especially at the earlier latencies between approximately $0\,\mathrm{ms}$ and $20\,\mathrm{ms}$, the brain response of the musicians is stronger than the brain response of the non-musicians. To test the significance of the differences be-

tween the magnitudes of the two groups, t-tests were again conducted at the same latencies as for the comparison of the attended and ignored mode on the population level. At $8\,\mathrm{ms}$, the thereby obtained p-value is $p = 0.94$, at $27\,\mathrm{ms}$ it is $p = 0.44$ and at $53\,\mathrm{ms}$ it was found to be $p = 0.82$. Therefore, the differences between the strenghts of the neural responses of the two groups at these three latencies are not significant. Especially for $8\,\mathrm{ms}$, this result seems surprising, considering the visual impression of the magnitude plot. This issue is further illuminated in section 5.4.

The magnitudes of the musicians and the non-musicians in the ignored mode of the fundamental waveform in Figure 5.3 **B** predominantly overlap as well for the latencies from approximately $30\,\mathrm{ms}$ on. At latencies smaller than that, again, the neural response of the musicians appears to be stronger than the neural response of the non-musicians. By conducting t-tests however, there was no significant difference found. The p-values turned out to be $p = 0.15$ at $8\,\mathrm{ms}$, $p = 0.61$ at $27\,\mathrm{ms}$ and $p = 0.90$ at $53\,\mathrm{ms}$.

The magnitudes of the TRFs of the envelope modulation of the musicians and the non-musicians in the attended mode in Figure 5.3 **C** largely overlap for the entire latency interval. Visually, there are no greater differences between the brain responses of the two groups, except for some peaks between approximately $5\,\mathrm{ms}$ and $30\,\mathrm{ms}$. The conducted t-tests confirmed this impression. The p-values were found to be $p = 0.98$ at $5\,\mathrm{ms}$, $p = 0.99$ at $23\,\mathrm{ms}$ and $p = 1.00$ at $47\,\mathrm{ms}$. With all of these being far larger than 5%, there is no significant difference between the strengths of the neural responses of the musicians and the non-musicians at these latencies.

Lastly, in Figure 5.3 **D**, the magnitudes of the envelope modulation of the musicians and the non-musicians in the ignored mode have a very similar shape, although the musicians seem to show stronger responses overall. For all of the peaks, the brain responses of the musicians visually show higher magnitudes than those of the non-musicians. Again however, the t-tests turned out to deem the differences insignificant for all three considered latencies. The p-values resulted in $p = 0.27$ at $5\,\mathrm{ms}$, $p = 0.36$ at $23\,\mathrm{ms}$ and $p = 0.43$ at $47\,\mathrm{ms}$.

5.3 Influence of musical training on the attentional modulation of the brain response

To investigate whether musical training influences the attentional modulation of the brain responses, the magnitudes of the TRFs of the attended and ignored mode were compared within the groups of musicians and non-musicians for both acoustic features, respectively. The four comparison plots are depicted in Figure 5.4. The magnitudes were computed with the second approach described in section 4.5 and section 5.2. Again, t-tests were performed, according to section 4.6, at the same latencies as in the sections before, which are marked by the dashed lines. Note that the population-averaged plots were obtained through taking the average across subjects and then the magnitude of the result-

ing TRF, while for the statistical analysis a distribution was obtained through taking the magnitude value of each single-subject. Therefore, the here presented p-values may seem not matching with the depicted population-averaged magnitude plots. A detailed discussion of this averaging issue is provided in section 5.4 and subsection 6.1.4.

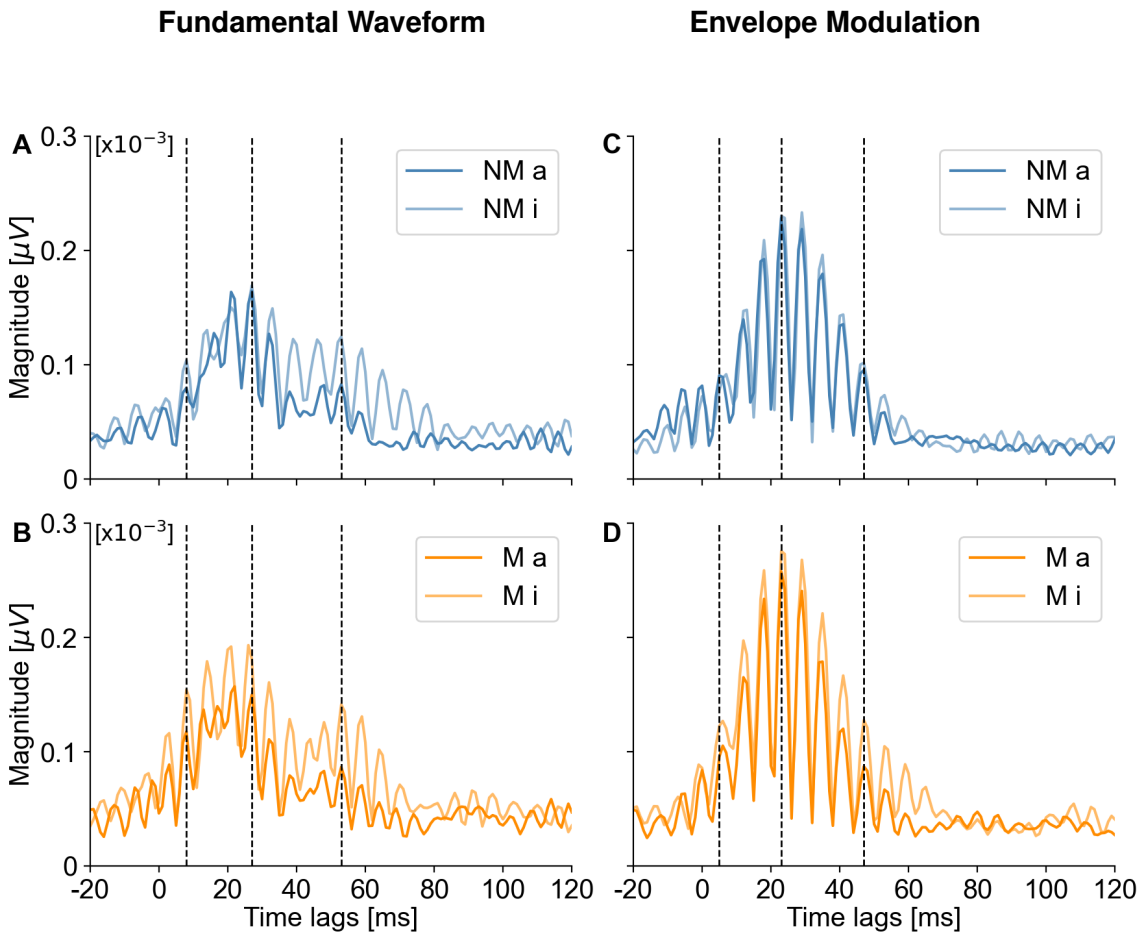


Figure 5.4: Influence of musical training on the attentional modulation of the brain response. Magnitude comparison of the attentional modes withing the groups of the musicians (**B**,**D**) and non-musicians (**A**,**C**) for the fundamental waveform (**A**,**B**) and the envelope modulation (**C**,**D**). The dashed vertical lines mark the latencies at which the t-tests were performed. Within the non-musicians, significant differences were found at the latest considered latency for both acoustic features. Within the musicians, significant differences were found for the subcortical brain response and at the latest considered latency for both acoustic features.

Figure 5.4 **A** compares the neural responses of the non-musicians to the fundamental waveform in the attended and ignored mode. Visually, there seems to be no difference between the strength of the magnitudes until approximately 35 ms. Then, until about 80 ms, the brain response above noise level is stronger in the ignored mode than in the attended mode. The conducted t-tests supplied p-values of

$p = 0.39$ at $8\,\mathrm{ms}$, $p = 0.08$ at $27\,\mathrm{ms}$ and $p = 0.01$ at $53\,\mathrm{ms}$. Therefore, the difference in the strength of the neural responses of the non-musicians in the attended and ignored mode can be deemed significant at 53 and insignificant $8\,\mathrm{ms}$ an $27\,\mathrm{ms}$.

The comparison of the magnitudes in the attended and ignored mode of the fundamental waveform of the musicians in Figure 5.4 **B** shows overall greater magnitudes for the ignored response within the interval of neural response above noise level (approximately $0\,\mathrm{ms}$ to $70\,\mathrm{ms}$). The results of the t-tests mostly validate this observation. The computed p-values are $p = 0.05$ at $8\,\mathrm{ms}$, $p = 0.06$ at $27\,\mathrm{ms}$ and $p = 0.02$ at $53\,\mathrm{ms}$. Thus, the differences between the brain responses of the musicians in the two attentional modes are significant at $8\,\mathrm{ms}$ and $53\,\mathrm{ms}$ and just slightly above the significance threshold of 5% at $27\,\mathrm{ms}$.

In Figure 5.4 **C**, the brain responses in the attended and ignored mode of the envelope modulation of the non-musicians are shown. Over the entire latency interval of response above noise level, the magnitudes of the two attentional modes largely overlap. The p-values that resulted from the t-tests were found to be $p = 0.46$ at $5\,\mathrm{ms}$, $p = 0.80$ at $23\,\mathrm{ms}$ and $p = 0.02$ at $47\,\mathrm{ms}$. This implies, that there is no significant difference between the two magnitude curves at the first two considered latencies. Although it seems unlikely by observation of the plot, there is a significant difference between the responses of the non-musicians in the attended and ignored mode at $47\,\mathrm{ms}$. The reason behind this mismatch of apparent visual and statistical result is further explained in section 5.4.

Figure 5.4 **D** shows the magnitudes of the TRFs of the envelope modulation in the attended and ignored mode of the musicians. The two magnitude curves have very similar shapes, with the one of the ignored mode having slightly greater magnitude values at the peaks of the brain responses. With the t-tests, a significant difference was found between the two attentional modes of the musicians, with a p-value of $p = 0.03$ at $5\,\mathrm{ms}$ and with $p = 0.01$ at $47\,\mathrm{ms}$. At $23\,\mathrm{ms}$, the difference between the strength of the neural responses of the ignored and attended mode can not be deemed significant, due to a p-value of $p = 0.21$.

5.4 Influence of the computational order on the magnitudes

In this section, the interesting change of the magnitudes of the TRFs, depending on which of the averaging orders depicted in section 4.5 was chosen, is illustrated. The starting point therefore was the plot comparing the magnitudes of the non-musicians and musicians in the attended mode of the fundamental waveform in Figure 5.3 **A**. This plot is also depicted in Figure 5.6 **C**. Although the neural response of the musicians at $8\,\mathrm{ms}$ visually seems to be stronger than the response of the non-musicians, the t-test deemed the difference insignificant, with an obtained p-value of $p = 0.94$. Since this seemed to be a mismatch of the visually observed and the statistical results, the origin therefore needed to be investigated.

The assumption was, that the process of computing the magnitudes caused this mismatch. For the t-tests, the magnitude values of the single-subject TRFs are used (see section 4.6). In the order of calculation that was used to obtain the population-averaged magnitudes shown in Figure 5.6 **C** however, these single-subject magnitudes practically never appeared in the computation process. But they are specifically used in the first approach to the magnitude computation introduced in section 4.5. There, the absolute values of the channels of the single-subject TRFs are taken, followed by the channel average over all 64 EEG channels. The result from these two steps is exactly what is used in the t-tests. Only afterwards, the population average is computed. Therefore, it was assumed that the results of the t-tests might be reflected better by the magnitudes that were obtained using the first explained process in section 4.5. Thus, the results and how they were altered by changing the averaging order in the magnitude computation needed to be examined and compared. To picture the differences of the two processes of magnitude computation more clearly, they are schematically depicted in Figure 5.5.

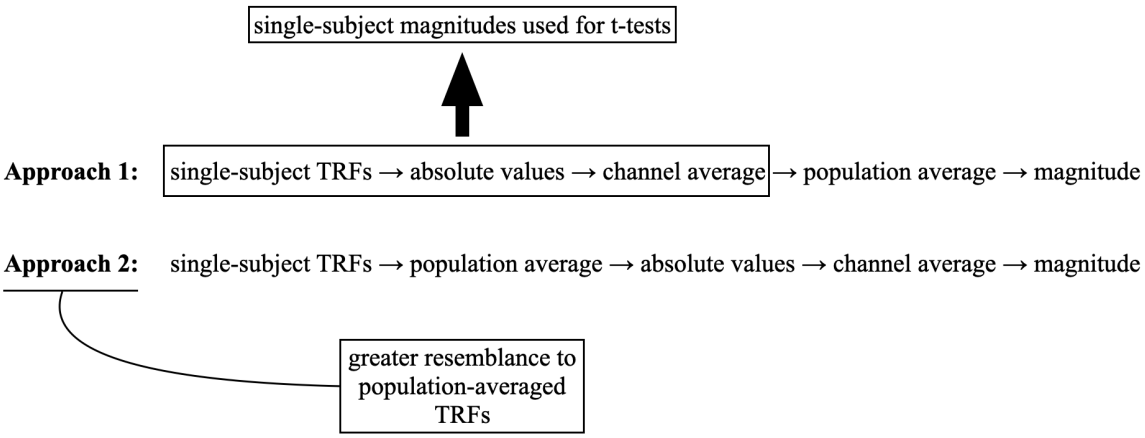


Figure 5.5: Schematic depiction of the two different approaches to the magnitude computation. With approach 1, first, the single-subject magnitudes are computed, which are also used for the t-tests. Approach 2 was chosen for the analysis of the data in this bachelor thesis, as it provides a greater resemblance to the population-averaged TRFs.

Figure 5.6 **A** shows the comparison of the magnitudes of the fundamental waveform of the non-musicians when the lower pitch speaker was attended, computed with the first approach (darker blue) and the second approach (lighter blue). Although both magnitude curves are still in the same scale, the first-approach magnitude seems to be positively vertically shifted compared to the second-approach magnitude. The general shape of the two magnitudes remains the same. The latency interval of neural response in both cases lies between approximately $0\,\mathrm{ms}$ and $60\,\mathrm{ms}$. Figure 5.6 **B** shows the comparison of the magnitudes of the fundamental waveform of the musicians when the lower pitch speaker was attended, computed with the first approach (brown) and the second approach (orange). Again, the first-approach magnitude is shifted vertically compared to the second-approach

magnitude, although the strength of the responses relative to their respective noise level seems to be approximately the same. Now the latency interval of neural response above the noise level appears to be broader for the second-approach magnitude, than for the first-approach magnitude. The second-approach magnitude shows earlier responses, or at least stronger responses at smaller latencies (approximately $0\,\mathrm{ms}$ to $15\,\mathrm{ms}$).

Figure 5.6 **C** shows, as mentioned before, the comparison of the magnitudes of the TRFs of the fundamental waveform of the musicians and the non-musicians, while the lower pitch speaker was attended. These magnitudes were computed with the second approach described in section 4.5. This plot is the same as Figure 5.3 **A** and has already been analysed in section 5.2. In Figure 5.6 **D**, the magnitudes of the TRFs of the fundamental waveform of the musicians and the non-musicians in the attended mode were computed with the first approach depicted in section 4.5. The magnitudes in the plots Figure 5.6 **C** and **D** originate from the same dataset and were even derived from the same TRFs, for the musicians and non-musicians respectively. However, due to the population average being calculated at different steps of the magnitude computation, they do not turn out to be equal. The reason behind this discrepancy is explained and discussed in detail in subsection 6.1.4. The three dashed vertical lines in both plots again mark the latencies, at which the t-tests were performed. As already stated in section 5.2, the p-values turned out to be $p = 0.94$ at $8\,\mathrm{ms}$, $p = 0.44$ at $27\,\mathrm{ms}$ and $p = 0.82$ at $53\,\mathrm{ms}$. While these appear to not reflect the visual difference between the magnitudes in Figure 5.6 **C**, especially at $8\,\mathrm{ms}$, the p-values seem to be sensible when comparing them to the magnitudes presented in Figure 5.6 **D**. Here, the magnitudes of the musicians and the non-musicians lie very close to each other and nearly completely overlap at the three chosen latencies.

In the end, considering the better comparability of the magnitudes obtained by the first approach to the t-test results, it was still decided to use the second approach for the magnitude computation for its greater resemblance to the TRFs. This decision will be discussed further in section 6.1.

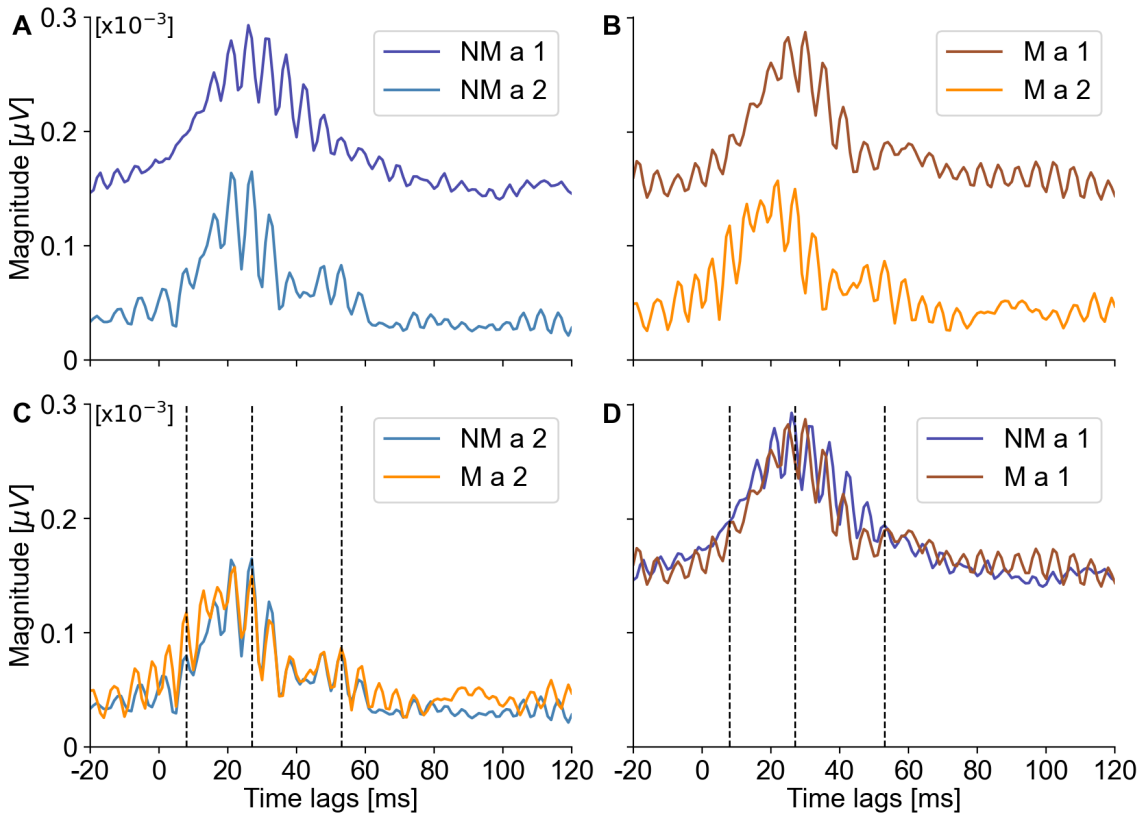


Figure 5.6: Influence of different computational orders on the magnitudes. Comparison of the two versions of averaging employed on the data (attended mode - fundamental waveform) of the **A** non-musicians and **B** musicians. Comparison of the magnitudes of the musicians and non-musicians computed with averaging method **C** 2 and **D** 1. The dashed vertical lines mark the latencies at which the t-tests were performed in the sections above. No significant differences were found at these latency times between the magnitudes of the musicians and non-musicians. While this seems surprising in comparison to plot **C**, it seems sensible when observing plot **D**.

6 Discussion

As the previous chapter presented the results obtained by the methods described in chapter 3 and chapter 4 used on the provided EEG data, this chapter intends to classify and interpret the various findings described in chapter 5. To do so, this chapter is divided into two parts. In the first section, the results are classified in regard to their meaning in the research field of neuroscience. The second section discusses some of the methods that were specifically tried and used in this bachelor thesis (chapter 4) and their impact on the results, as well as some possibilities they provide for future work.

6.1 Results

6.1.1 Measurement of subcortical and cortical brain responses

As briefly mentioned before, EEG is most commonly used to capture the earlier subcortical brain responses, like the auditory brainstem response [11, 12]. MEG on the other hand, is predominantly used to measure cortical neural responses, that occur later than the subcortical responses [28, 26, 6]. The EEG data used for this bachelor thesis provided TRFs and topoplots, as seen in Figure 5.1, that hint to the measurement of both subcortical and cortical neural responses. Furthermore, the later cortical responses are even stronger than the earlier subcortical responses. This holds for the fundamental waveform and the envelope modulation, as well as for the population average of all 50 participants, the musicians and the non-musicians. As illuminated in section 2.3, this is however not unusual, since EEG measures both subcortical and cortical neural activity. Since the electrodes are closer to the cortex than to the subcortical area of the brain, it is also not surprising, that the cortical responses are stronger than the subcortical responses.

Additionally, responses at even greater latencies could be captured, which however can not be definitely classified as either subcortical or cortical responses, since the corresponding topoplots show the area of neural activation between the regions that are typical for either of these two types of responses. An assumption would be, that these responses reflect the top-down neural communication from the cortex back to the subcortex. This is however only a very vague idea of interpretation, for which up to now no robust research data can be found, and should therefore be seen critically.

6.1.2 Stronger neural responses in ignored mode than in attended mode

In section 5.1, the attentional modulation of the neural responses is presented, on the population level, as well as for some single subjects. For the population-averaged responses to the fundamental waveform, a significant difference was found at some depicted latencies between the mode, where the participants attended the lower pitch speaker, and the mode, where they ignored him. Contrary to previous studies on this topic however, the neural response was stronger in the ignored mode, than in the attended mode for all three considered latencies. Generally however, the TRFs and corresponding magnitudes of the two attentional modes appear to be very similar and to overlap for a majority of the observed latencies. This result also applies to the brain response to the envelope modulation as acoustic feature. The work of Riegel [26] on the corresponding MEG data of 43 of the here considered subjects found, that the cortical neural responses were significantly stronger when attending the lower pitch speaker, than when ignoring him. Also, Schüller et al. [28] observed significantly stronger cortical brain responses on the population level, when the lower pitch speaker was attended, than when he was ignored. They used the MEG data of 22 subjects, the simultaneously recorded EEG data of 16 of them was used in this bachelor thesis. So, compared to both of these studies, a similar result was originally expected for the analysis of the here used EEG data, since it was recorded at the same time as the MEG data for mostly the same subjects. These two studies using MEG focused on the cortical responses. The studies of Etard et al. [11] and Forte et al. [12] investigated the influence of attention on the auditory brainstem response, so on the earlier subcortical responses, that were also captured in the data used for this bachelor thesis. They both found, that the subcortical response to the attended mode was significantly stronger than the response to the ignored mode, also contradicting the findings of this bachelor thesis.

By investigating the influence of attention on the neural response within the groups of musicians and non-musicians in section 5.3, similar results were observed. For the responses of the non-musicians to the fundamental waveform, no significant difference was found between the subcortical and cortical responses to the ignored and attended mode of the lower pitch speaker. For the responses at the greatest chosen investigated latency ($53\,\mathrm{ms}$), the neural response to the ignored mode was significantly stronger than to the attended mode. A similar observation was made for the brain responses of the musicians, however here, the subcortical response could just merely be deemed significant with a p-value of $p = 0.05$. Not one of the two groups by themselves seems to cause the unusual result of the ignored mode leading to stronger responses than the attended mode. However, it should be noted, that the p-values of the musicians were far closer to the threshold of suggesting a significance (cortical $p = 0.06$) than those of the non-musicians (subcortical $p = 0.39$, cortical $p = 0.08$). In regard to the envelope modulation, within the group of the non-musicians, the differences between the attended and ignored mode were only significant at $47\,\mathrm{ms}$, where the response when ignoring the lower pitch speaker was again stronger. Within the musicians, the ignored subcortical response and the one at $47\,\mathrm{ms}$ were significantly stronger. Finally, these results could suggest that musicality has a larger impact on the ignored (subcortical) responses, as these were significantly stronger than the

attended responses on the population level. Musicians often have to listen to different melodies and instruments simultaneously, for example when playing as part of an orchestra. This could potentially lead to stronger responses to background noises compared to non-musicians, as they are trained to concentrate on these simultaneously to the main noise. However, this would have to be investigated further, e.g. by using the membership in a band or orchestra as a distinguishing parameter within the group of the musicians. It is still rather questionable, that the responses in the ignored mode are generally stronger than those in the attended mode.

The unusual finding, that the ignored mode mostly leads to stronger neural responses than the attended mode, can be explained mathematically by the observation of the single-subject TRFs and magnitude plots of all 50 subjects. In doing so, it was found, that the data of most subjects resembled the 'unexpected' data of participant 2, which was presented in Figure 5.2 **B** and **E**. So, for most subjects, the data showed a stronger magnitude of the ignored responses than of the attended responses. Only a few subjects resembled the 'expected' participant 52, who's data is presented in Figure 5.2 **C** and **F**. Therefore, it is sensible, that the total population average, as well as the averages within the groups of musicians and non-musicians, mirror the results of the majority of the considered subjects.

It is important to note, that the quality of the individual results seems to be independent of the group affiliation. Considering the data presented in Figure 5.2 again, subject 13 is neither a musician, nor a non-musician. Subject 2 and 52 are both non-musicians. Nevertheless, by observing the magnitudes of the single-subject TRFs of each participant, there seems to be no correlation between the quality of the results and whether the subjects count as musician, non-musician or neither.

The reason for these results, that are contradictory to previous studies on this topic, could be the use of passive electrodes for the EEG measurement of the data used in this bachelor thesis. Active electrodes are the most commonly used option for EEG measurements, when no simultaneous different measurements, like the MEG in this case, are performed. For instance, Etard et al. [11] used active electrodes in their study on the attentional modulation of the EEG measured auditory brainstem response. On the other hand, Forte et al. [12] also used passive EEG electrodes and found, that the subcortical responses were stronger when attending the narrator than when ignoring him. For now, it remains unclear, how these rather unexpected results of this bachelor thesis came to be. For future measurements with the used experimental setup, it would be important to first investigate, whether these unusual findings could be explained by the equipment or setup in general. Finally, it is important to note, that speech processing is still only poorly understood. It leaves room for more research in the future, also regarding the attentional modulation of the neural responses.

6.1.3 No significant difference between musicians and non-musicians

The data provided no significant differences between the neural responses of the musicians and the non-musicians in the competing speaker scenario. This applies for the earlier subcortical responses, as well as for the cortical responses. All of the conducted t-test concluded insignificant differences between the musicians and non-musicians when comparing the attended responses to the fundamental waveform, the ignored responses to the fundamental waveform, the attended responses to the envelope modulation and the ignored response to the envelope modulation. While the magnitude plots in Figure 5.3 seem to suggest otherwise, section 5.4 explains the reason for the mismatch of visual impression from the magnitude plots and the statistical results, which will also be discussed further in the following subsection.

These results were not expected, especially regarding previous findings on this topic, that suggest that musicians generally show better auditory skills than non-musicians [16, 22, 31, 34], such as faster or stronger neural responses, for instance. Even in these studies however, the differences between musicians and non-musicians were only small and should therefore be seen critically. Furthermore, the insignificant differences between the two groups in this case align with a very recent development within the research group. Riegel found significant differences in the cortical contributions to the speech-FFRs between musicians and non-musicians, by analysing the MEG data that was measured simultaneously to the here used EEG data in [26]. Meanwhile however, the data of more subjects was observed and included in the evaluation. This led to the results aligning with the findings of this bachelor thesis. The study that presents these newer outcomes is still in the process of being published.

Additionally it should be noted, that in the process of this work, the topoplots of the musicians and non-musicians were compared for different latencies. Contrary to the non-musicians, the musicians seemed to clearly show a subcortical neural activation to the fundamental waveform ($8\,\mathrm{ms}$). This was not specifically included and discussed in this bachelor thesis however, since it is only a vague result. Nevertheless, it was an interesting observation that could be investigated further.

While the aspect of musicality was intended to be the main focus of the analysis, it provided no significant results. It is important to note however, that the general quality of the EEG data, as discussed in the subsection above, should also be taken into account for the assessment of this issue. Furthermore, the criteria for determining whether a participant qualifies as a musician or a non-musicians could be chosen more stringently in future work on this topic. This could possibly lead to a greater difference in the results of these two groups as well.

6.1.4 Influence of magnitude computation order

As previously mentioned in this discussion and presented in section 5.4, an interesting difference was found between the magnitudes of the TRFs, depending on the computational order that was chosen. While the first approach (absolute values → channel average → population average) provided a better comparability of the p-values obtained by the conducted t-tests, the second approach (population average → absolute values → channel average) showed a greater resemblance to the original population averaged TRFs (see also Figure 5.5). It is important to note, that the averaging order to compute the magnitudes does not change the t-test results, as these were conducted using the magnitudes of the single-subject TRFs. Therefore, the statistical statements resulting from the analysis of the speech-FFRs are distinct. Within the research group, the common way of magnitude computation is the first approach [28, 26].

With this in mind, the second approach was chosen to be the used for the magnitude computation nevertheless, since the effects found in the TRFs were better depicted with this computational order. Especially for the comparison of the neural responses to the fundamental waveform of the musicians and non-musicians (see Figure 5.3), we wanted to display the earlier rise in the magnitude curves of the musicians. These were not visible in the magnitude plots computed with the first approach.

These differences are still important to keep in mind for future work, for a better understanding and classification of results represented by their magnitudes. The origin of the differences between the two approaches could not be clearly determined. One assumption is, that the different averaging orders lead to different interferences of the signals of the EEG channels. For the total population average with the second approach for example, 50 waveforms of one channel are added up, which certainly leads to interferences of the signals. These specific interferences do not occur within the first approach, as the absolute values of the single-subject TRFs are calculated first. It seems reasonable, that this leads to different outcomes. For future work, it would be interesting to further investigate this observation, to find out how the magnitudes are exactly altered by the different computation steps.

The cause of the vertical shift in Figure 5.6 between the magnitudes computed with the first and second approach remains rather unclear. The above mentioned interferences should not change the baseline values of the signals. An explanation could be, that the single-subject TRFs, or at least some of the EEG channels, show negative or positive offsets. In the first approach, these are then also transferred to the single-subject magnitudes in the form of only positive offsets, due to the calculation of the absolute values of each channel and the following channel average. In the second approach, the positive and negative offsets are averaged out, since the population average of the single-subject TRFs is computed first. Lastly however, the vertical shift does not change the relative magnitude values of the brain responses.

6.1.5 Influence of chosen latency on statistical results

Lastly, the statistical results represented by the p-values obtained from the t-tests should be approached critically. This holds for the significant, as well as the insignificant outcomes. Independent of the magnitude computation order, it can be observed that there are rather strong fluctuations in the magnitude curves. Therefore, choosing different latencies than those that were used, could lead to different statistical results. Regarding the comparison of the attended response to the fundamental waveform of the musicians and non-musicians computed with the first approach in Figure 5.6 **D**, for instance. Choosing $5\,\mathrm{ms}$ or $11\,\mathrm{ms}$ instead of the here considered $8\,\mathrm{ms}$ could possibly suggest a significant difference between the subcortical responses of the musicians and the non-musicians. This was not done however, since we wanted to conduct the t-tests at latencies, where neural responses were suggested by the TRFs and corresponding topoplots. This is another reason why it would be important to further look into the alterations of the magnitudes depending on the computation approach. If the TRFs suggest certain latencies that would be interesting for the t-tests, these latencies should then also actually be sensible for the t-tests. Furthermore, taking a range of latency times, according to the duration of the neural response into account, instead of only single latencies, the statistical results would also be more reliable. If the differences between the attended and ignored mode, as well as between the musicians and the non-musicians were more drastic in general, this problem would not occur anyway. For this bachelor thesis however, this critical approach to the statistical results should be kept in mind.

6.2 Methods

6.2.1 The alignment check

The alignment check introduced in section 4.4 provides an improvement of the preprocessing workflow for the analysis of the EEG data. Through this process, it was found that the software that is usually used within the research group to temporally align MEG and EEG data is lacking in precision. Furthermore, it was possible to also correct the slight temporal shift between the aligned MEG and EEG data later on. This improved the results, since the latencies of the single-subject TRFs now better represented the actual time delay between the presentation of the acoustic stimulus and the neural response of the participant. The improvement on the single-subject level then also led to more expressive TRFs on the population level and within the musicians and non-musicians.

The alignment check is rather time consuming, since it has to be done for each participant separately and involves some manual input as well. The execution of the alignment check approximately doubles the temporal effort that has to be put into the alignment of the MEG and EEG data in the first place. Thus, a further improvement for similar work in the future could be to either generally improve the precision of the alignment software, or to implement the alignment check in a way that is more automated, maybe even as a separate function within the original software. Since I am not familiar with the development of the software, it is rather difficult to say to which extent any of those ideas could be realized.

Finally, the subject-specific slight imperfection after the first alignment of the EEG and MEG datasets, which was discovered with the alignment check, is an important outcome of this bachelor thesis. It should be noted, that this only had a larger impact on this work, since a temporal shift of approximately $1\,\mathrm{ms} - 5\,\mathrm{ms}$ significantly changes the timing of the brain responses depicted by the TRFs that occur between $5\,\mathrm{ms}$ and $60\,\mathrm{ms}$. Since the software has thus far been used mostly for the analysis of neural responses at greater latencies, this lack of precision was not apparent.

6.2.2 Reproducibility of results with other acoustic features

As presented in section 4.2, two different acoustic features than the fundamental waveform and the envelope modulation were tried as well, but discarded in the end. These were the low-frequency envelope and the high-frequency envelope of the audiobooks narrated by the lower pitch speaker. Mainly due to the comparability to the results found by the analysis of the MEG data [26] that corresponds to the here used EEG data, the fundamental waveform and the envelope modulation were finally chosen to be used as the acoustic features.

Although, or maybe especially because, the results found in this bachelor thesis are not as meaningful as hoped for, it would be interesting to further investigate the EEG data with different acoustic

features. The low-frequency envelope was only evaluated on the total population average. It would thus be interesting to analyse the EEG data with this acoustic feature with regards to the differences between musicians and non-musicians. This could either give a greater insight into the general quality and reliability of the EEG data, or even provide results that suggest significant differences in the neural responses between those two groups. Since there was not enough time to additionally cover this within this bachelor thesis, it provides an interesting task for future research.

The TRFs and magnitudes derived with the high-frequency envelope as the acoustic feature did not deliver meaningless results, but rather results that were not more insightful than by using the fundamental waveform and the envelope modulation. Since the high-frequency envelope introduced in [33] is still a rather new acoustic feature, it also provides attractive opportunities for further research.

6.2.3 Possibility of individual regularization parameters

Alternatively to using the optimal regularization parameter of $\lambda = 5$ for the TRFs of all considered participants, it was also tried to work with individual regularization parameters (see section 4.3). Unfortunately, I could not find any research on a similar topic that used this approach. Therefore, and since it did not seem to provide vastly different results than by using the same λ for the data of all subjects, this idea was discarded.

It could be an interesting topic for future research however. The use of individual regularization parameters would provide a superior regularization for the data of each participant. Therefore, the lastly resulting TRFs could potentially better reflect the strength and timing of the actual neural responses. This potential improvement or decline in the quality of the results could be illuminated, by testing if the difference in using one collective or individual regularization parameters is significant.

7 Conclusion and Outlook

The aim for this bachelor thesis was to analyse the EEG-measured speech-FFRs of musicians and non-musicians. To do so, the EEG data of 50 participants was used in total. During the EEG measurements, the subjects simultaneously listened to two audiobooks, that were narrated by two speakers with slightly different pitch, while having to focus their attention on only one of them. In doing so, it was intended to investigate whether their attendance would have an impact on the strength of their neural responses to the presented acoustic stimulus.

The EEG data was preprocessed and temporally aligned with the corresponding MEG data, that was recorded simultaneously, to finally be temporally aligned with the respective audiobook narrated by the lower pitch speaker. During this temporal alignment, a lack in precision of the software, that was used for this step, could be determined by implementing an alignment check. With this alignment check, it was possible to correct the slight temporal shift between the EEG and MEG datasets, and therefore between the EEG data and the audiobooks, for each of the participants. This correction led to an improvement of the temporal classification of the neural responses.

The EEG data was then used, together with two acoustic features of the audiobooks narrated by the lower pitch speaker, to compute Temporal Response Functions (TRFs). The chosen acoustic features were the fundamental waveform and the envelope modulation of the higher modes of the fundamental frequency of the lower pitch narrator. Other acoustic features were tried as well, namely the low- and high-frequency envelopes of the audiobooks narrated by the lower pitch speaker. These were however discarded, primarily for the comparability to previous work within the research group. Then, the magnitudes of the TRFs were computed. Two different approaches were therefore tried. While the first approach is commonly used within the research group, the second approach was chosen, due to its greater resemblance to the shape of the TRFs. Further investigation showed, that the first approach however, better depicts the statistical results that were found by conducting t-tests at specific latencies. This difference in the outcome of the magnitudes, depending on the chosen computational order, is an interesting finding of this bachelor thesis.

Using the data of all 50 participants, for certain latencies, a significant difference was found between the strengths of the neural responses when the participants were attending and ignoring the lower pitch narrator. This was observed for the subcortical, cortical and even later response to the fundamental waveform and for the cortical and later response to the envelope modulation. At the observed latencies, the ignored response was stronger than the attended response, which was not expected and does not align with earlier research on this topic as discussed in section 6.1. For the most part how-

ever, the brain responses in the two attentional modes appeared to be similar.

Three criteria (see section 3.2) were used to divide the 50 participants into groups of musicians and non-musicians. In the end, 18 subjects could be categorised as musicians, 23 as non-musicians and the remaining nine subjects did not qualify for either group. Comparing the neural responses of the musicians and non-musicians, no significant differences could be determined for any attentional mode and any acoustic feature. By investigating the attentional modulation of the brain responses within the groups of musicians and non-musicians, some significant differences could be found. Within the non-musicians, the ignored responses at the greatest tested latency were significantly stronger than the attended responses, for both the fundamental waveform and the envelope modulation. Within the musicians, the subcortical responses and the late responses to the fundamental waveform and the envelope modulation were stronger in the ignored than in the attended mode.

In conclusion, no significant differences were found by the comparison of the musicians and the non-musicians, and rather unusual and surprising results were found by considering the attentional modulation of the neural responses. However, as previously discussed, the statistical results should be approached critically, since they are strongly dependant on the exact latencies at which the t-test were conducted. It is important to note, that human speech processing is still only poorly understood. This leaves more questions for future work within this field of research. On the other hand, improvements could be made to the general data processing methods used for this bachelor thesis, by implementing an alignment check to correct the slight temporal shift of the EEG data. Lastly, the difference that was found between the magnitudes, depending on the order of their computation, is an interesting finding that should be taken into account for future work.

Interesting issues to approach in future research could be the later neural responses, which were captured here at $53\,\mathrm{ms}$ and $47\,\mathrm{ms}$, especially focusing on their origin in the brain and their meaning to human speech processing. Also, further investigation of the influence of musicality on auditory skills and the processing of continuous speech could provide a better understanding of this topic. A discrimination within the musically trained subjects could be implemented, for example by differentiating between the played instruments or whether the participants play as parts of music groups, such as bands or orchestras. Additionally, further work could be considered on the differences between using passive or active electrodes for EEG measurements, and how they impact the quality of the results, respectively. Furthermore, it would be interesting to investigate the possibility of using subject-specific regularization parameters for the TRF computation process. Lastly, the differences in the magnitudes, depending on the chosen computational order, could be studied further.

Bibliography

[1] Andrea Biasiucci, Benedetta Franceschiello, and Micah M. Murray. "Electroencephalography". In: Current Biology 29.3 (2019), pp. 80–85. DOI: https://doi.org/10.1016/j.cub.2018.11.052.

[2] Gavin M. Bidelman. "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR". In: Hearing Research 323 (2015), pp. 68–80. DOI: https://doi.org/10.1016/j.heares.2015.01.011.

[3] Ralf Brandes, Florian Lang, and Robert F. Schmidt. Physiologie des Menschen. mit Pathophysiologie. Springer Berlin, Heidelberg, 2019. URL: https://doi.org/10.1007/978-3-662-56468-4.

[4] Christian Brodbeck and Jonathan Z. Simon. "Continuous speech processing". In: Current Opinionin Physiology 18 (2020), pp. 25–31. DOI: https://doi.org/10.1016/j.cophys.2020.07.014.

[5] Taishih Chi, Powen Ru, and Shihab A. Shamma. "Multiresolution spectrotemporal analysis of complex sounds". In: The Journal of the Acoustical Society of America 118.2 (2005), pp. 887–906. DOI: 10.1121/1.1945807.

[6] Emily B. J. Coffey, Sibylle C. Herholz, Alexander M. P. Chepesiuk, Sylvain Baillet, and Robert J. Zatorre. "Cortical contributions to the auditory frequency-following response revealed by MEG". In: Nature Communications 7.11070 (2016). DOI: https://doi.org/10.1038/ncomms11070.

[7] National Institute on Deafness an other Communication Disorders. "How Do We Hear? (visited on Jan.31, 2024)". In: online (2022). URL: https://www.nidcd.nih.gov/health/how-do-we-hear.

[8] Wolfgang Demtröder. Experimentalphysik 1. Springer Spektrum Berlin, Heidelberg, 2017. DOI: https://doi.org/10.1007/978-3-662-54847-9.

[9] Giovanni M. Di Liberto, James A. O'Sullivan, and Edmund C. Lalor. "Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing". In: Current Biology 25 (2015), pp. 2457–2465. DOI: https://doi.org/10.1016/j.cub.2015.08.030.

[10] Bradley Efron and Trevor Hastie.
Computer age statistical inference: Algorithms, evidence, and data science.
Cambridge University Press, 2016. DOI: https://doi.org/10.1017/CBO9781316576533.

[11] Octave Etard, Mikolaj Kegler, Chananel Braimanb, Antonio Elia Fortea, and Tobias Reichenbach. "Decoding of selective attention to continuous speech from the human auditory brainstem response". In: NeuroImage 200 (2019), pp. 1–11.
DOI: https://doi.org/10.1016/j.neuroimage.2019.06.029.

[12] Antonio Elia Forte, Octave Etard, and Tobias Reichenbach. "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention".
In: eLife 6 (2017). DOI: https://doi.org/10.7554/eLife.27203.

[13] Selenia di Fronso, Patrique Fiedler, Gabriella Tamburro, Jens Haueisen, Maurizio Bertollo, and Silvia Comani. "Dry EEG in Sports Sciences: A Fast and Reliable Tool to Assess Individual Alpha Peak Frequency Changes Induced by Physical Effort".
In: Frontiers in Neuroscience 13 (2019). DOI: 10.3389/fnins.2019.00982.

[14] Alexandre Gramfort et al. "MEG and EEG Data Analysis with MNE-Python".
In: Frontiers in Neuroscience 7.267 (2013), pp. 1–13. DOI: 10.3389/fnins.2013.00267.

[15] Mikolaj Kegler, Hugo Weissbart, and Tobias Reichenbach.
"The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information". In: Frontiers in Neuroscience 16 (2022).
DOI: https://doi.org/10.3389/fnins.2022.915744.

[16] Nina Kraus and Bharath Chandrasekaran.
"Music training for the development of auditory skills".
In: Nature Reviews Neuroscience 11 (2010), pp. 599–605.
DOI: https://doi.org/10.1038/nrn2882.

[17] Joshua P. Kulasingham, Christian Brodbeck, Alessandro Presacco, Stefanie E. Kuchinsky, Samira Anderson, and Jonathan Z. Simon.
"High gamma cortical processing of continuous speech in younger and older listeners".
In: NeuroImage 222 (2020).
DOI: https://doi.org/10.1016/j.neuroimage.2020.117291.

[18] Tao Li, Yitao Ma, and Tetsuo Endoh. "Neuromorphic processor-oriented hybrid Q-format multiplication with adaptive quantization for tiny YOLO3".
In: Neural Computing and Applications 35 (2023), pp. 1–29.
DOI: 10.1007/s00521-023-08280-y.

[19] Fabien Lotte, Laurent Bougrain, and Maureen Clerc.
"Electroencephalography (EEG)-Based Brain-Computer Interfaces".
In: Wiley Encyclopedia of Electrical and Electronics Engineering (2015).
DOI: 10.1002/047134608X.W8278.

[20] Matthias Mauch and Simon Dixon.
"PYIN: A fundamental frequency estimator using probabilistic threshold distributions". In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014, pp. 659–663. DOI: 10.1109/ICASSP.2014.6853678.

[21] Volker Milnik. "Anleitung zur Elektrodenplatzierung des internationalen 10–20-Systems". In: Das Neurophysiologie-Labor 31.1 (2009), pp. 1–35.
URL: https://www.sciencedirect.com/science/article/pii/S1439484708000781.

[22] Christo Pantev, Robert Oostenveld, Bernhard Ross, Manfried Hoke, Almut Engelien, and Larry E. Roberts. "Increased auditory cortical representation in musicians".
In: Nature 392 (1998), pp. 811–814. DOI: https://doi.org/10.1038/33918.

[23] Alexandra Parbery-Clark, Erika Skoe, Carrie Lam, and Nina Kraus.
"Musician Enhancement for Speech-In-Noise".
In: Ear and Hearing 30.6 (2009), pp. 653–661. DOI: 10.1097/AUD.0b013e3181b412e9.

[24] M. P Paulraj, Kamalraj Subramaniam, Sazali Bin Yaccob, Abdul H. Bin Adom, and C. R Hema. "Auditory Evoked Potential Response and Hearing Loss: A Review".
In: Open Biomed Eng J. 9 (2015), pp. 17–24. DOI: 10.2174/1874120701509010017.

[25] Diana C. Peterson, Vamsi Reddy, Marjorie V. Launico, and Renee N. Hamel.
"Neuroanatomy, Auditory Pathway." In: StatPearls [online] (updated 2023).
URL: https://www.ncbi.nlm.nih.gov/books/NBK532311/#.

[26] Jasmin Riegel. Analysis of Speech-FFRs in Musicians and Non-Musicians. 2023.

[27] Achim Schilling et al. "Analysis of continuous neuronal activity evoked by natural speech with computational corpus linguistics methods".
In: Language, Cognition and Neuroscience 36 (2020), pp. 1–20.
DOI: 10.1080/23273798.2020.1803375.

[28] Alina Schüller, Achim Schilling, Patrick Krauss, Stefan Rampp, and Tobias Reichenbach. "Attentional modulation of the cortical contribution to the frequency-following response evoked by continuous speech". In: Journal Of Neuroscience 43 (2023a), pp. 7429–7440.
DOI: https://doi.org/10.1523/JNEUROSCI.1247-23.2023.

[29] Alina Schüller, Achim Schilling, Patrick Krauss, and Tobias Reichenbach. "Early subcortical response at the fundamental frequency of continuous speech measured with MEG".
In: bioRxiv (2023b). DOI: https://doi.org/10.1101/2023.06.23.546296.

[30] "Sound, N. (3)". In: Oxford English Dictionary, Oxford UP, December 2023.
URL: https://doi.org/10.1093/OED/876234983.

[31] Dana L. Strait, Nina Kraus, Erika Skoe, and Richard Ashley. "Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion".
In: The European journal of neuroscience 29.3 (2009), pp. 661–668.
DOI: https://doi.org/10.1111/j.1460-9568.2009.06617.x.

[32] Gregor Strobbe. “Advanced forward models for EEG source imaging”. PhD thesis. Jan. 2015. URL: https://www.researchgate.net/publication/294430587_Advanced_forward_models_for_EEG_source_imaging.

[33] Mike Thornton, Jonas Auernheimer, Constantin Jehn, Danilo Mandic, and Tobias Reichenbach. “Detecting Gamma-Band Responses to the Speech Envelope for the ICASSP 2024 Auditory EEG Decoding Signal Processing Grand Challenge”. In: in print (2024).

[34] Donald Watanabe, Tal Savion-Lemieux, and Virginia B. Penhune. “The effect of early musical training on adult motor performance: evidence for a sensitive period in motor learning”. In: Exp Brain Res 176 (2007), pp. 332–340. DOI: 10.1007/s00221-006-0619-z.

[35] Dezhong Yao, Yun Qin, Shiang Hu, Li Dong, Maria L. Bringas Vega, and Pedro A. Valdéz Sosa. “Which Reference Should We Use for EEG and ERP practice?” In: Brain Topography 32 (2019), pp. 530–549. DOI: https://doi.org/10.1007/s10548-019-00707-x.

Eigenständigkeitserklärung

Hiermit versichere ich, Melanie Wolf (22792002), die vorgelegte Arbeit selbstständig und ohne unzulässige Hilfe Dritter sowie ohne die Hinzuziehung nicht offengelegter und insbesondere nicht zugelassener Hilfsmittel angefertigt zu haben. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen und wurde auch von keiner anderen Prüfungsbehörde bereits als Teil einer Prüfung angenommen.

Die Stellen der Arbeit, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angaben der Herkunft kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet.

Verstöße gegen die o.g. Regeln sind als Täuschung bzw. Täuschungsversuch zu qualifizieren und führen zu einer Bewertung der Prüfung mit „nicht bestanden“.

Ort, Datum	Melanie Wolf