Brainstem response estimation using continuous sound

A feasibility study

Julia Adlercreutz

LUND UNIVERSITY

Department of Automatic Control

MSc Thesis
TFRT-6177
ISSN 0280-5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

Printed in Sweden by Tryckeriet i E-huset
Lund 2022

Abstract

Hearing loss is a complicated phenomena which does not only vary from person to person, but also, can change characteristics during the day. Despite this, hearing aids today are fitted only occasionally and thus only capture the slow changes in the hearing loss. In order for a hearing aid to continuously adapt to a subject's hearing loss it has to be able to gauge the users hearing threshold. One way of measuring the hearing threshold is by examining the auditory brainstem response (ABR).

The problem with measuring the ABR today is that it has to be measured as the response to a short sound that is repeated thousands of times. This masters thesis investigates a new method of estimating the brainstem's response to continuous sound. This new paradigm builds on the assumption that the brainstem response corresponds to an impulse response to a system that takes the heard audio as input, and gives the EEG recording as output.

This thesis explores how well this new paradigm works on hearing impaired people that use hearing aids. It verifies that the method works for finding the impulse responses that resemble the cortical response, which is a stronger and slower response. The method was however not successful when it came to finding the subcortical response. A possible reason for this is that a lot of the data needed to be removed due to outliers.

Acknowledgements

Thank you to all my supervisors; Emina Alickovic, Martin Skoglund and Hamish Innes-Brown from Eriksholm Research Center and Bo Bernhardsson from the Department of Automatic Control at Lund University. They have provided a wide range of knowledge and help me with everything from understanding the brain and the auditory pathway to finding problems in my code. They have come with many great ideas during our weekly meetings and gave me a lot of feedback that helped shape the project. Their input truly has been invaluable.

Also a big thank you to my examiner Maria Sandsten from Mathematical Statistics at Lund University for hear feedback and for the discussion of the project.

Thank you to my parents Dietlind and Patrick Adlercreutz and my sister Becca. Your love and support means a lot to me and I will be forever grateful to have you in my life.

Also a huge thank you to Erik Prändtare who besides supporting me also brought my computer back to life after it crashed in the final month of the project. I can not thank you enough for that.

Contents

1

Introduction

The processing of sound and speech is a complicated task that involves multiple different parts of the brain. The sound travels through the ear canal and the inner ear until it reaches the cochlea, where the vibrations of the sound trigger an electrical impulse that propagates through the auditory nerve, brainstem, and thalamus to the cortex. While the cortex is the part of the brain that is usually studied with regard to speech processing, the electrical impulse actually goes through significant processing before reaching the cortex[Maddox and Lee, 2018]. It is the subcortical activity, and more specifically, the sound processing in the brainstem, that is studied in this thesis.

To gain an understanding of how speech is being processed by the brain, brain activity has to be measured. One method to do this is with electroencephalography (EEG). The EEG measurements are typically recorded with electrodes spread out over the scalp. The components of the EEG signal that correspond to a stimulus response is small compared to the rest of the EEG signal [Luck and Kappenman, 2012], which in this case, can be seen as noise. This means that the responses have a low signal-to-noise ratio (SNR). Cortical activity has higher SNR than subcortical, partly explaining why cortical activity is more commonly studied [Maddox and Lee, 2018]. Furthermore, subcortical responses, such as the brainstem response, have a high-frequency content [Luck and Kappenman, 2012], making these signals hard to distinguish from noise.

Because of the low SNR, the brainstem response is usually studied through evoked potentials (EP). This means exposing a person to a repetitive stimulus while recording the EEG signals. This repetitive stimulus can for example be a chessboard pattern that switches colors, or, as is the case when the auditory response is studied, a series of click sounds. The EEG signal is then split into smaller segments, here called epochs. Each epoch is associated with one instance of the repetitive stimulus, that is to say, one color change of the checkered pattern or one click sound. The average EEG signal is then calculated for all the epochs. This average is the

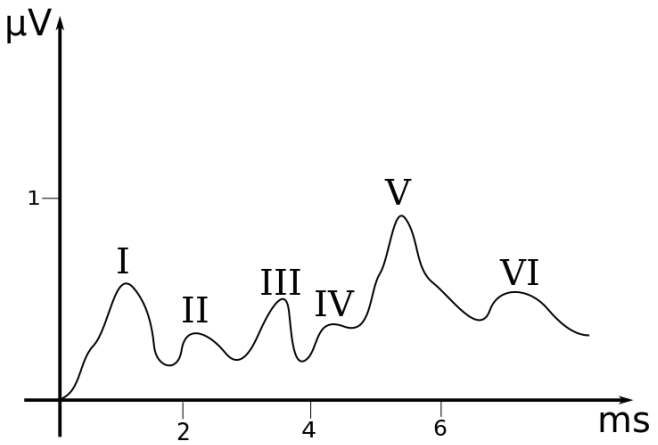


Figure 1.1 A schematic figure showing what a traditional ABR can look like. The strongest peak is called wave V and appears at average 5.7 ms after the sound is played.

EP and shows the brain's response to the stimulus. The EP reflects both cortical and subcortical activity [Laguna and Sörnmo, 2005]. This method of obtaining the brainstem response is based on signal detection theory. It is based on the assumption that the response is stationary and embedded in noise that is uncorrelated with it. Since the noise is uncorrelated it is different from epoch to epoch and the averaging will therefore uncover the stationary brainstem response [David M. Green, 1966].

The EP that is used to study the subcortical processing of sound, is called the auditory brainstem response (ABR) and is usually obtained when a test subject is listening to clicks, although other short sounds can also be used. A sketch of what such a response can look like can be seen in Figure 1.1.

The ABR consists of seven waves which are numbered with roman numerals according to the order they appear in the responses. These waves are sometimes studied in clinical settings since the morphology and latencies of the wave carry valuable information about a patient's hearing. For example, the larger peaks, such as wave V, can be used to determine hearing thresholds [Lv et al., 2007]. This is for example used with infants when examining their hearing [WHO, 2010].

There is a great limitation, however, with the traditional ABR. Since it is typically obtained by averaging over thousands of short sounds, it can not be used to investigate the subcortical processing of natural speech. Instead, the stimulus has to be

repetitive which might lead to some patients losing focus. The unnatural sounds might also be distressing to listen to for some patient groups like those with dementia. Furthermore, the limitations in the stimulus greatly affect the ability to do research about subcortical sound processing. For instance, it is not possible to investigate whether the brainstem responds the same way to its native language and an unknown language when the sound has to be so short.

Looking at the ABR from a mathematical modeling point of view, it might seem natural to view the click sound as an impulse of stimulus. The ABR, in turn, could be seen as a system response to an impulse of stimulus. In other words, the ABR would act as an impulse response to the brainstem. If this would be the case, algorithms from system identification could be used to obtain the ABR. The possible advantage of computing the ABR this way is that there exist algorithms to identify impulse responses even if the stimulus is not an impulse.

This way of estimating impulse responses and comparing them to cortex related EPs has been done in a number of studies, for example [Lalor et al., 2009] and [Aljarboa et al., 2022]. Recently the method has also been adapted to find impulse responses that are thought to correspond to the ABR [Maddox and Lee, 2018]. This impulse response was called the temporal response function (TRF) or the speech-derived ABR, since it was estimated with continuous speech as stimulus. It carries morphological similarities to the traditional ABR, for example, it has a peak corresponding to wave V in the traditional ABR.

It should be noted that despite there being promising morphological similarities between the impulse responses and their corresponding EPs, they are not necessarily equivalent. For one, the impulse response relies on the assumption that the system is linear and time-invariant, which is not necessarily true. Furthermore, it has been suggested that the impulse response does not always perform the same as its corresponding EP and that it can not be used in the same clinical setting for diagnosis [Lalor et al., 2008].

As mentioned before, the traditional ABR can be used to determine hearing thresholds. This begs the question of whether the speech-derived ABR can be used in the same way. If the speech-derived ABR can be used in the same way, it would allow for the determination of the hearing threshold using natural sounds. This in turn opens up for the possibility of hearing devices that continuously can measure hearing thresholds and automatically adjust the hearing aid setting if need be.

So far, however, the research conducted has only had normal hearing test subjects and it is therefore not clear how and if a hearing impairment affects the speech-derived ABR. This thesis aims to give more insight into this by estimating the speech-derived ABR methods using data collected from hearing-impaired test sub-

jects. Furthermore, the data used in this report has been collected under other conditions than previously. Instead of using insert phones to play the sound stimulus, the sound was here played from a loudspeaker. This adds more uncertain transfer steps into the analysis which might complicate the system identification.

The main goal of this thesis is to explore whether it is possible to find an impulse response resembling the ABR using this new data set. It also sets out to find suitable ways of evaluating the results.

2

Background

2.1 EEG signals

This section gives a brief overview of EEG signals, what they are and what problems that can occur when working with them. Most of the information stated is inspired by [Laguna and Sörnmo, 2005].

Electroencephalography (EEG) is a method of recording brain activity. As the name suggests, the method works by recording the electrical activity in the brain that stems from the neurons in the brain, which, like other neurons, send signals that are partially electric.

The EEG is usually measured by placing a number of electrodes on the scalp. Sometimes they can also be placed directly on the cortex, measuring the non-invasive scalp potential, however, is much more common. The scalp electrodes are usually placed according to the international 10/20 system. [Mitra and Bokil, 2008]

In the 10/20 system, each electrode has a two or three character long name. The first is a letter that describes the region of the scalp; F for frontal, P for parietal, C for central, T for temporal, O for occipital, and A for auricle. Sometimes two letters are used when more electrodes are placed in between the standard electrodes. The second character describes the angle with which the electrode deviates from the central line that goes from the nose to the neck. This second character is typically a number, but it is a z (for zero) for electrodes on the central line. Figure 2.1 shows the placement of the different electrodes.

The electrodes record electrical potentials, this means that a reference potential is needed. Preferably, the reference should be affected by the same noise as the measurement electrodes without being affected by the neural activity. Common ways of constructing a reference include, averaging the potential from all electrodes and

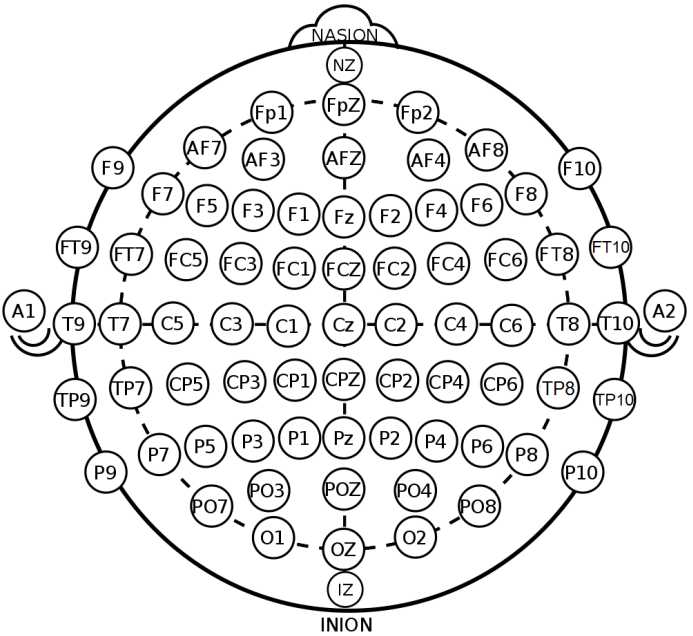


Figure 2.1 How electrodes can be placed during an EEG measurement and what they are named. Image by [Oxley, 2020].

using an electrode on the mastoid, which is a location behind the ear. [Mitra and Bokil, 2008]

The EEG signal is oscillatory. In general, if a person is active during the measurement the oscillation will have a higher frequency content than if the person was resting. The shape of the EEG spectrum typically follows a 1/f curve, meaning that lower frequencies are stronger.

Artifacts and noise

There is a lot of noise and artifacts in a raw EEG signal, both of a technical and biological origin. This section aims to give insight into common sources of noise, how to identify it, and, if possible, what to do about it.

Noise with technical origin. A common noise source in any type of electrical measurement is the power line. Depending on the country, this noise is either a 50 Hz or a 60 Hz sinusoid. It is the oscillating current in the power line that has the possibility of inducing noise in the electrodes. This can be mitigated by shielding the measurement setup from the power line as well as other electronic devices that produce the noise. The noise can easily be identified when looking at the frequency content of the EEG signal through a so called periodogram, an estimation of the

frequency content. Typically, clear peaks will be visible at 50/60 Hz if there is power line interference. Multiples of the base frequency can appear also.

If the recording contains power line interference, it can be dealt with using signal processing. The simplest way is to use a filter which targets and stops one specific frequency. Such a filter is called a notch filter and is described in further detail in Section 2.4.

The next type of technical noise is a low frequency drift in the EEG signal. This can be taken care of with a high-pass filter, which removes the slow changes in the base line.

The base line can also be affected when the electrode moves slightly. This is called an electrode-pop, and as the name suggest, it results in a sudden change in the base line.

Noise with biological origin. When measuring the EEG, the only interesting signals are those associated with brain activity. Unfortunately, many processes in the body give rise to electrical fields that are indistinguishable from the electrical activity to the recording electrode. It is therefore important to use knowledge about the different electrical signals when analysing EEG signals.

One of the common sources of biological artifacts are eye-movements and eye-blinks. These can be seen as large bumps in the EEG, that is to say they have a fairly low frequency content. These eye-related movements alter the electric potential between the cornea and the retina and are more pronounced the closer the electrode is to the eyes. A common way of removing these artifacts from the EEG signal is by measuring the eye-related electrical fields separately with a so called electrooculogram (EOG). The EOG signal can then be used in artifact cancellation algorithms. If an EOG signal is not available, however, the simplest way to handle the artifacts is to identify them and simply exclude them from the modeling or analysis.

Similarly to eye-movements, the electrical activity of the heart can also be measured in the EEG. These artifacts also have a low frequency content but usually have a small amplitude compared to the EEG. For some people and measurements, however, the heart interference might be quit evident. They are most easily recognized by the fact that they have a very regular occurrence in the signal with one bump for every heart beat. This activity of the heart can be measured with an electrocardiogram (ECG) which can also be used in artifact cancellation algorithm. Alternatively, samples that contain heart activity can be excluded when doing the analysis.

If many EEG channels have been recorded, but neither EOG or ECG, then indepen-

dent component analysis (ICA) principle component analysis (PCA) can be used [Mitra and Bokil, 2008]. ICA, as the name suggests, separates mixed signals into independent components. Both the EEG and the EOG are independent from the EEG but mixed into the recording. These separated out independent components can be used to cancel the noise, similarly to when the EOG and ECG had been recorded. Similarly, PCA identifies the principle components, which also can correspond to the noise sources [Hyvärinen et al., 2001].

A draw-back with using PCA or ICA is that it requires multiple electrodes. The number of independent or principle components that can be found with the algorithms, is the same as the number of signals that are used in the algorithm. This means that if only three EEG channels have been recorded, three independent or principle components can be found, even if there exist more underlying components [Hyvärinen et al., 2001].

Lastly, muscle movements in the face can give rise to noise as well. This is typically a noise in the range of 15-30 Hz. If these frequencies are not of interest for the analysis, a filter can be used to remove these frequencies.

2.2 Rectified signals

In the context of this thesis, a rectifier is a digital preprocessing step that ensures that the processed signal is non-negative. Two different methods for rectifying signals are mentioned in this thesis, although more methods exist.

The first method is called half-wave rectification and is illustrated in Figure 2.2. In this method, negative values are replaced with zeros.

The second method is called full-wave rectification and is also illustrated in Figure 2.2. Instead of replacing the negative samples with zero, their signs is flipped, resulting in positive values.

2.3 Auditory evoked potential (AEP) and event related responses

The traditional auditory evoked potentials

Evoked potentials (EPs), as mentioned in chapter 1, are responses in the EEG signal that are associated with an event, e.g. a click sound. EPs have a comparatively low amplitude, only about one tenth of the background EEG or less, and are therefore hard to observe directly. The EPs are therefore estimated from EEG signals that have been recorded for a long time, with many identical events, e.g. thousands of

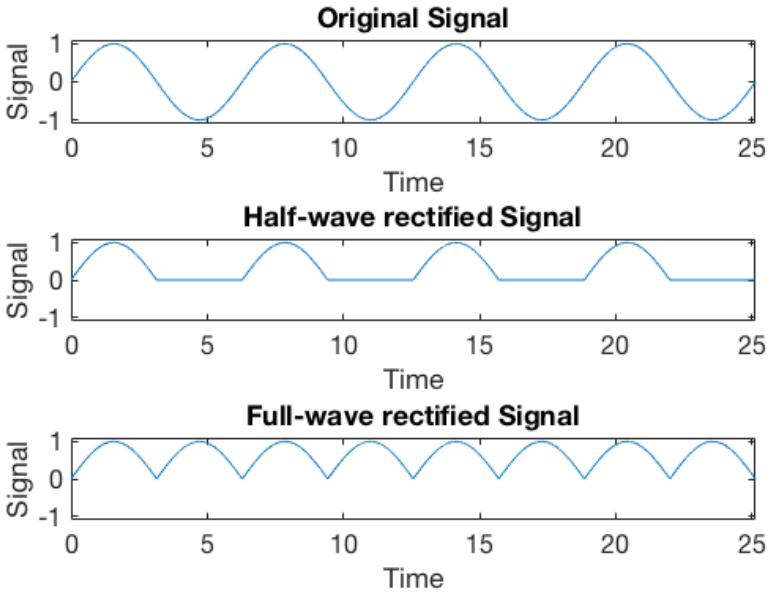


Figure 2.2 An example of different rectifying methods, i.e. different methods of making a signal non-negative. The top most plot shows the signal before it has been rectified. The middle plot shows a half-rectified signal and the bottom one a full-wave rectified signal.

click sounds. The EEG signal is then divided into segments, called epochs, were each epoch is associated with one instance of the event. The EP is then obtained as the average of EEG across all epochs. [Laguna and Sörnmo, 2005]

The auditory evoked potential is divided into three different parts, depending on latency. The earliest responses is the auditory nerve and brainstem response (ABR), which span the latencies up until 10 ms after the event. After that follows the middle-latency evoked potentials from 10 to 60 ms, and lastly the long latency components at latencies 60 to 200 ms. In this thesis. The earliest responses have a much smaller amplitude. This is because the response stems from more and more neurons as the responses get further along the auditory pathway. More neurons do not only lead to a stronger signal but also one with a lower frequency content since the neurons are not as well synchronised [Luck and Kappenman, 2012].

The ABR is typically measured with an electrode on top of the head in the Fz or Cz location in the 10/20 system, that is to say, on the forehead, and a reference electrode close to the ear, e.g. on the mastoid or the ear lobe. The response consist of five to seven fast oscillations, with peaks that are denoted with roman numerals

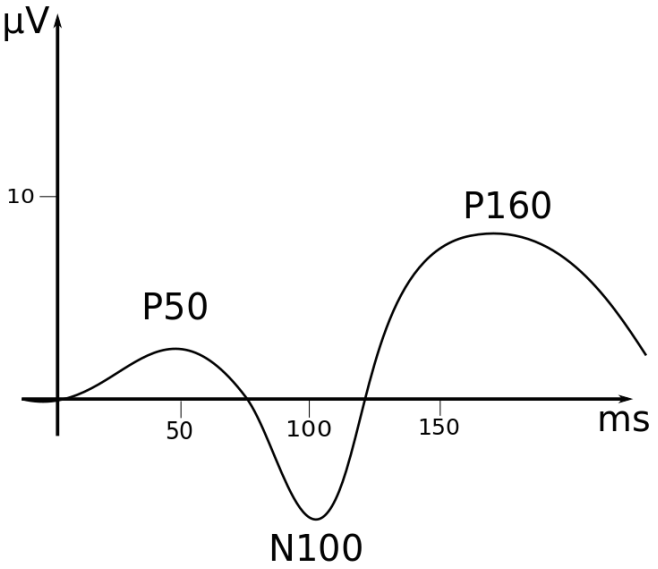


Figure 2.3 A schematic figure showing what a traditional long-latency response to sound can look like. It consists of three clear peaks at 50, 100 and, 160 ms.

in order of latency. The strongest peak is wave V with and it has a latency of about 6 ms. [Luck and Kappenman, 2012] A schematic plot of what the ABR can look like can be seen in Figure 1.1. The ABR response is non-linear with respect to the stimulus intensity. The response does not only have a higher amplitude, but is also faster when the stimulus is stronger [Nousak and Stapells, 2009].

The first peaks in the long-latency are denoted P50, N100, P160, and N200. The letters P and N are used to signal that the peak is positive and negative respectively. The number indicates at what latency the peak is expected in milliseconds, meaning that P50 is a positive peak that occurs 50 ms after the stimulus. [Luck and Kappenman, 2012] A schematic plot of what the long-latency response can look like can be seen in Figure 2.3.

Impulse responses and evoked potentials

The first comparisons between AEPs and impulse responses were done in [Lalor et al., 2009]. In that study, the new event-related response was called auditory-evoked spread spectrum (AESPA) and it was inspired by similar work done with visual EPs [Lalor et al., 2006]. While the AEP is a traditional EP obtained by averaging, the AESPA method is based on the assumption that the EEG signal is the output

from the convolution of the amplitude modulated audio signal and an unknown impulse response. It is this unknown impulse response that is the AESPA and it was estimated linear least-squares, using regularisation to reduce the over all estimation error.

The sound signals used during the first AESPA experiments were broadband Gaussian noise, pure tones and a mixture of broad band Gaussian noise and short sounds used when measuring a traditional AEP. These AESPA responses were compared to a regular AEP obtained when averaging the EEG signal from when the stimulus is a short repetitive sound. The results showed morphological similarities between the AEP and the AESPA with peaks in the group average AESPA that correspond to peaks in the AEP. The found peaks all corresponded to cortical activity, meaning that the AESPA did not show the response of the brain stem. The paper [Lalor et al., 2009] further shows that the AESPA can give rise to the same response as the AEP when generated with discrete stimuli and that the AEP is strongly spatially correlated to the AESPA when it is generated with a tone. That is to say that the two responses come from more or less the same parts of the brain. The paper suggests that the AEP and the AESPA stem from similar neural generators in the brain and that the AESPA is a generalization of the AEP.

The AESPA was later expanded to use speech as stimulus [Lalor and Foxe, 2010]. In this new AESPA, the speech envelope is used as the input to the system, i.e. the speech envelope was used as a regressor in the least-squares estimation of the impulse response. The envelope of a speech signal only captures the rough outline of how the signal varies, but no details. This was argued to be acceptable since it has been shown that it is the envelope that is the most important component in speech recognition. This second paper also suggests that the AESPA shows the response of cortical activity, and more specifically the response of the early auditory cortex.

Based on the work with the AESPA, efforts were made to capture the brainstem response to natural speech. In a paper by [Maddox and Lee, 2018] the brainstem response was estimated using linear regression, however, this time a half-wave rectified audio signal was used as a regressor, not the speech envelope. This new regressor has the advantage of preserving high frequency components which are essential for the ABR. The signal was rectified to reflect the fact that the auditory pathway is not affected by the sign of the incoming sound.

In [Maddox and Lee, 2018], the EEG recordings were done with three different types of stimuli; a pseudorandomly timed click-train, a periodic click-train and continuous speech. The traditional ABR was also calculated as the response to the periodic click-train, and it was used to validate the response of the pseudorandom timed click-train. The pseudorandomly click-train, in turn, was compared with the speech-derived response. The responses were compared using Pearson's correlation

coefficient and the results show that the response to the two click trains is highly similar, meaning that the pseudorandom clicks produce a response which is similar to periodic clicks. The pseudorandom click train was also shown to have strong similarities with the speech-derived response. The speech-derived response also has peaks, just like the ABR, wave V being the peak that showed up most consistently across the responses although wave VI was also visible in the average response.

2.4 Filtering

Filtering is a common step when preprocessing EEG signals. High-pass filters can be used to remove base-line drift in the signal, notch filters can be used to remove power-line interference and band-pass filters can be used when analysing a specific frequency band in the EEG. However, one has to be careful when filtering since the filters usually affect the signals more than just in the desired way [de Cheveigné and Nelken, 2019].

In this section, it is described what is meant by a filter, different attributes that are used to describe a filters properties, as well as what pitfalls one have to be aware of when filtering.

General information about filters

Filtering can be done with hardware in continuous time. Most of the filtering in this thesis, however, happens in the digital domain. In this case, filtering a signal $x(n)$ with a filter, with impulse response $h(n)$, results in a signal $y(n)$ which is computed

$$y(n) = \sum_{i=-\infty}^{\infty} h(i)x(n-i) = h(n) * x(n) \tag{2.1}$$

were $*$ denotes the convolution operator, in this case, discrete convolution. The letter n is used to indicated that the signals exist in the discrete time domain.

There are four major types of filters; low-pass, high-pass, band-pass and band-stop filters. These describe how the filter affects different frequencies, e.g. a low-pass filter lets low frequencies pass through it while high frequencies are stopped by it. Band-pass and band-stop filters have a band of frequencies that are unaffected or stopped, respectively. A notch filter is a special case of a band-stop filter, were the band of the stopped frequencies is so thin that the filter more or less targets one specific frequency.

What constitutes a high frequency is described by the cutoff frequency, usually denoted f_c. An ideal low-pass filter, for instance, is supposed to not affect frequencies below the cutoff frequency at all, while all frequencies over the cut off should be

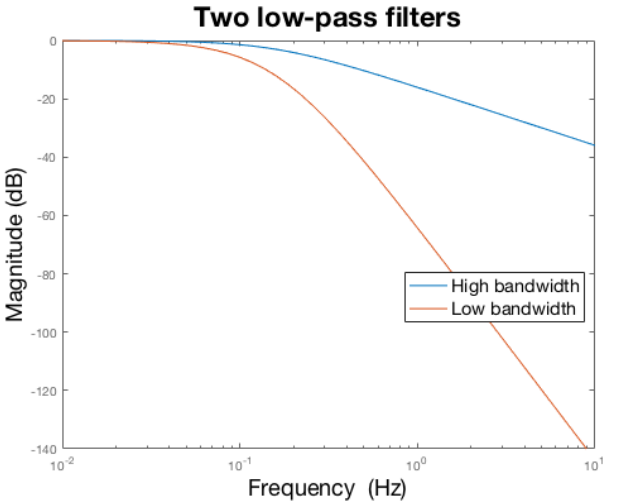


Figure 2.4 An example of two low-pass filters with the same cutoff frequency but with different bandwidths.

completely blocked. Band-pass and band-stop filters have two cut off frequencies, one for when the band starts and one for when it stops.

In reality, however, there is often a more gradual change between how frequencies are affected, and there will be a band of frequencies that are attenuated by the filter without being stopped completely. This is illustrated in Figure 2.4, which also shows how this band of frequencies can be different between filters even if the cutoff frequency is the same. There are multiple ways of describing this band of frequencies, in this thesis, the bandwidth will be used. It can be defined slightly differently, but here it is defined as the frequency at which the attenuation has reached -3 dB. The bandwidth, in other words, describes the steepness of the slopes in Figure 2.4 and is different for the two different features.

Another way to characterise a filter is to describe whether it has a finite impulse response (FIR) or an infinite impulse response (IIR). As the names suggested, these descriptions characterise the impulse response $h(n)$ and if it has a finite or infinite length.

Problems that arise from filtering

As mentioned before, filters can be a wonderful tool that helps when analysing signals, however, they can also affect the signals in undesirable ways. This section describes some of the problems that filters can cause and how to handle them.

Causality. One aspect to consider when using a filter is that of causality. A causal filter is if the output signal only depends on past values of the input. In terms of the impulse response function, $h(n)$ from equation (2.1), this means that $h(n) = 0$ for all negative n [Devasahayam, 2000]. Similarly an anti-causal filter only depends on future values of the input and only has non-zero elements for negative n. If a filter is neither causal nor anti-causal, it is referred to as an acausal filter.

To investigate the causality of a filter, one should study the filter's impulse response. If this is not readily available, it can be obtained by creating an impulse signal, that is a signal $\delta(n)$ such that

$$\delta(n) = \begin{cases} 1 & ; \quad n = 0 \\ 0 & ; \quad n \neq 0. \end{cases} \tag{2.2}$$

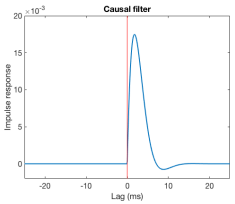
The impulse response can than simply be obtained by applying the filter to $\delta(n)$.

Upon investigation of the definition of a filter in equation (2.1), it might seem peculiar that you could apply a filter without having access to the impulse response. However, in reality it is common to use filtering functions in e.g. MATLAB, and in that case, the impulse response is typically hidden within the function and not known by the user of it.

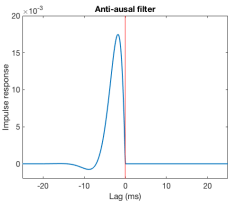
For an example of the different causalities, see Figure 2.5, were three different impulse responses with different causalities have been plotted. The base filter that has been used to produce the plots is a 2nd order Butterwort filter with cutoff frequency 100 Hz, which is causal. The anti-causal filter was obtained as suggested by [Maddox and Lee, 2018], by reversing the input signal, applying a causal filter, and then reversing the output of the filter. The acausal filter was done by applying the causal filter both in the forward and backward direction using MATLAB's function `filtfilt` (resulting in a fourth order filter).

An important thing to notice in Figure 2.5, is that the input has been spread out over time. Recalling equation (2.2), the input signal was only non-zero at lag 0, while the output signals are non-zero for multiple lags. The causal filter has been spread out forward in time, meaning that the output is active longer than the input signal was. The anti-causal filter, on the other hand, is non-zero before the input signal has become non-zero. And the acausal filter becomes non-zero before the input and stops being non-zero after it.

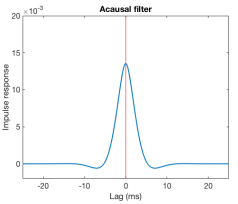
It is recommended in [Maddox and Lee, 2018] that, when preprocessing the signals to find the impulse response from audio signal to EEG signal, only to use causal filter on the EEG signal and anti-causal filters on the audio signal. If this principle is not followed, this leads to an acausal estimated impulse response. This would



(a) A causal filter which only depends on past values of the input signal because the impulse response is zero for negative lags.



(b) An anti-causal filter which only depends on future values of the input signal because the impulse response is zero for positive lags.



(c) An acausal filter which depends on both past and future values of the input signal because the impulse response is non-zero for both positive and negative lags.

Figure 2.5 Three different impulse responses with different causality, the vertical line indicates lag 0. The filter has had the effects of spreading out the input signal over multiple time lags.

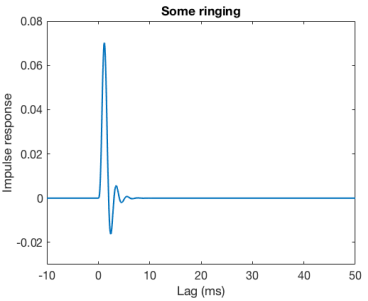
correspond to the EEG signal being affected by sounds that will be heard in the future, which should not be the case.

Another reason causality is important in this thesis, is because of the way the source of a response is determined. As mentioned previously, in a traditional AEP the earliest peaks, that occur just milliseconds after the stimulus, are said to correspond to subcortical activity. However, if wrong causality filters have been used, a peak that is associated with cortical activity can occur early in in the estimated response. If one is not cautious in this case, one might mistake cortical activity for subcortical activity.

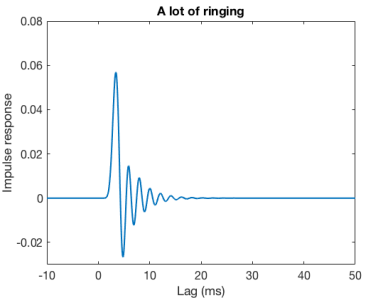
Another aspect to be aware of is that, similarly to how some peaks can occur earlier, peaks can also be delayed because of the filtering. This means that, even if filtering is done with the right causality, one should be careful when drawing conclusions that relay heavily on at which lag the peak occurs, if the delay caused by filtering has not been carefully examined.

Ringing. Some times, filtering a signals leads to oscillations in the output signal even if there were no oscillations in the input. This effect is called ringing and becomes more pronounced when the cutoff in the frequency domain is sharper. [de Cheveigné and Nelken, 2019]

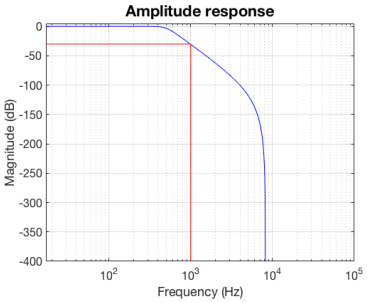
The oscillations tend to have a frequency close to the cutoff frequency [de Cheveigné and Nelken, 2019]. This information can help separate the ringing from other oscillations that occur naturally in the EEG and audio signals.



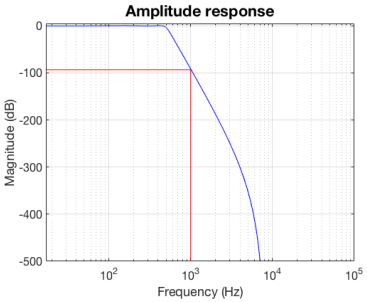
(a) An impulse response that shows some ringing.



(b) An impulse response that shows a lot of ringing.



(c) The amplitude response of the filter associated with Figure 2.6a.



(d) The amplitude response of the filter associated with Figure 2.6b.

Figure 2.6 A comparison between two filters with different amounts of ringing. It is evident that the filter with more ringing has a sharper cutoff in the frequency domain, exemplified by that the attenuation for $f = 1000$ Hz is about 50 dB stronger for the filter with a lot of ringing.

Just like with the question of causality, it is good to plot the impulse response to get a feeling for the ringing of the filter. It is also good to plot the amplitude response to see how sharp the cutoff in the frequency domain is. For an example of this, see Figure 2.6.

2.5 Statistics

The following section contains definitions of some statistical terms that are used in this thesis.

Periodogram

Let $x(n)$ be a signal vector of length N, where $n = 0, 1, 2...N-1$, and let $X(f)$ be the Fourier transform of the signal. Then the periodogram $\hat{R}_x(f)$ is

$$\hat{R}_x(f) = \frac{1}{n}\left|X(f)\right|^2 \tag{2.3}$$

and is an estimate of the spectral density of the signal [Lindgren et al., 2014]. The periodogram has a large variance.

Correlation

Pearson's correlation coefficient r for two stochastic variables x and y is obtained as

$$r(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}\frac{x_i-\mu_x}{\sigma_x}\frac{y_i-\mu_y}{\sigma_y}, \tag{2.4}$$

where N is the number of observations of the variables, μ_x and μ_y are their mean values, and σ_x and σ_y are their standard deviations. The coefficient measures the linear dependence between two stochastic variables [MATLAB, n.d.]

Null hypothesis and confidence intervals

This Section is based on [Blom et al., 2017].

Let X be the stochastic variable with a probability density function $P_X(k)$ for an outcome k. Typically, $P_X(k)$ will be given by a standard density function that depends on one or more unknown parameters. For instance, X could have a binomial distribution, in which case

$$P_X(k) = \sum_{j=0}^{k}\binom{n}{j}p^j(1-p)^{n-j}. \tag{2.5}$$

Here k is the number of times a certain outcome, with a probability p of occurring, has happened, given n trials and $P_X(k)$ denotes the probability that at most k occurances happen during n trials. If the probability p is unknown, a null hypothesis can be formulated that can either be supported or rejected with an observation. For example, let the null hypothesis H_0 be

$$H_0 \ : \ p = p_0, \tag{2.6}$$

that is to say the the initial guess is that the probability is p_0, for some value of p_0.

This null hypothesis can then be compared to an new alternative hypothesis called H_1, which usually takes one of three forms

$$H_1 : \begin{cases} p > p_0 \\ p < p_0 \\ p \neq p_0. \end{cases} \tag{2.7}$$

To dismiss the null hypothesis an observation x either has to be large enough or small enough to make the null hypothesis unlikely. Given a desired probability s of rejection of a correct null hypothesis, the limit a should be found such that

$$\begin{cases} P(X \geq a) \lesssim s \\ P(X \leq a) \lesssim s \\ P(X \leq a_{lower}) \lesssim \frac{s}{2} \quad \text{and} \quad P(X \geq a_{upper}) \lesssim \frac{s}{2}, \end{cases} \tag{2.8}$$

respectively, for each variant of H_1. The notation $a \lesssim b$ means that a is less than b, but still close to b. means that If the observation x falls within the limit, i.e. smaller than a, larger than a, and between a_{lower} and a_{upper}, respectively, the observation is said to lay within the $(1-s)$ confidence interval of H_0. For instance, if $s = 0.05$ the confidence interval is a 95 percent confidence interval. If the observation is outside the confidence interval, however, the null hypothesis is rejected in favour of the new hypothesis H_1. Since the probability of falsely rejecting a correct null hypothesis should be low, the probability s should be chosen to be a small value.

The parameter s is usually called the p-value and is usually dented with a p. This symbol s was chosen here in order to avoid confusion with the probability p in the example.

The example above follows a stochastic variable with a specific and known type of distribution, however, the method works as well if the distribution has only been estimated from observations.

2.6 Estimating impulse responses

Impulse responses, sometimes called temporal response functions (TRFs), are a way to describe the relation between input and output to a linear time-invariant (LTI) system, not just filter systems as in section 2.4. Given the input $u(t)$ and the impulse response $h(t)$, the output $y(t)$ can be obtained as

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = x(t) * h(t), \tag{2.9}$$

were the time parameter t is used to indicate continuous time, and $*$, in this case, denotes continuous convolution. In this thesis, the input signal is an audio signal and the output is the EEG signal that has been recorded when the audio signal was played.

The impulse response is connected to the frequency response function H by the Fourier transform $\mathscr{F}$

$$H(\omega) = \mathscr{F}\left(h(t)\right). \tag{2.10}$$

If a Fourier transform is applied to equation (2.1), this gives

$$Y(\omega) = H(\omega)U(\omega), \tag{2.11}$$

where $Y(\omega)$ and $U(\omega)$ are the Fourier transforms of $y(t)$ and $u(t)$, respectively. This means that the impulse response can be estimated by estimating the frequency response function which in turn can be estimated with

$$\hat{H}(\omega) = \frac{Y(\omega)}{U(\omega)}, \tag{2.12}$$

where the hat notation $\hat{}$ indicates that $\hat{H}(\omega)$ is an estimate of $H(\omega)$. The frequency content can easily be obtained with the fast Fourier transform (FFT).

There are some problems, however, with estimating the frequency response function with equation (2.12). One major problem is that the estimation is not robust because division with zero arises if $U(\omega) = 0$ for some ω. A simple solution is to add a small constant ε into the denominator

$$\hat{H}(\omega) = \frac{Y(\omega)}{U(\omega) + \varepsilon}. \tag{2.13}$$

This might still cause problems, however, since $U(\omega)$ could be equal to $-\varepsilon$ for some ω. Further alteration is therefore necessary for example using the more robust estimation

$$\hat{H}(\omega) = \frac{Y(\omega)U^*(\omega)}{U(\omega)U^*(\omega) + \varepsilon}, \tag{2.14}$$

where * denotes the complex conjugate, meaning that the denominator now is strictly positive for all ω if ε is positive. Also, note that if $\varepsilon = 0$ equation (2.14) gives exactly the same result as the original equation (2.12). This way the transfer function is estimated as a Tikhonov-regularized Wiener Filter in the frequency domain [Wikipedia, n.d.]

The term $U(\omega)U^*(\omega)$ is the periodogram of the input signal. Unfortunately, as mentioned in Section 2.5, has a large variance. Similarly, $Y(\omega)U^*(\omega)$ also has a

high variance. This means that the estimate $\hat{H}(\omega)$ can also be expected to have a high variance unfortunately.

The parameter ε works similarly to the regularisation parameter in other types of modeling schemes. To determine a suitable value for ε one might look at the periodogram of the input signal since $U(\omega)U^*(\omega)$ has to be significantly larger than ε to affect the estimation. Typically, the spectrum will show a higher amplitude content at low frequencies. This means that ε can be chosen such that only frequencies lower than a certain cutoff frequency affect the estimation.

3

The data

The data was collected by Eriksholm Research Center prior to the start of this thesis. The study was approved by the ethics committee for the capital region of Denmark (journal number H-1-2011-033). The study was conducted according to the Declaration of Helsinki, and all the participants signed a written consent prior to the experiment.

The study consisted of 11 subjects. The subjects are called TP01-TP11 through out this thesis. For one of these subjects, the exact age was not recorded, only that the participant was between 70 and 79 years. The rest of the subjects were of the ages 76-85 with a mean of 79.7.

All of the participants had some level of hearing loss. The mean and standard deviation of the subjects' hearing loss can be seen in the audiogram in Figure 3.1. It shows the hearing threshold for different frequencies, that is, how much a sound of a given frequency must be amplified for the subject to hear that frequency. During the experiment they used test hearing aids with frequency specific gain based on their audiogram. All additional hearing aid features were turned off during the experiment.

During the experiment, the participants sat in a sound proof room with a loudspeaker 1.65 m straight in front of them. They where presented with the audio book with five stories by H.C Andersen narrated by a male speaker. The stories were; *Den Flyvende Kuffert*, *Den Standhaftige Tinsoldat*, *Storkene 1*, *Storkene 2* and *Klokken*. Some of the stories were longer and were therefore split in two. This resulted in a total of 8 tracks, each about 6 minutes long. The splitting is illustrated in figure 3.2. All the audio was in Danish, which was the native language of all the participants.

The participants were asked to relax and listen to the story. They where not specifically asked to attend to the story and it was not tested whether they had been listening. While they were listening to the audio book, one track at the time, their EEG

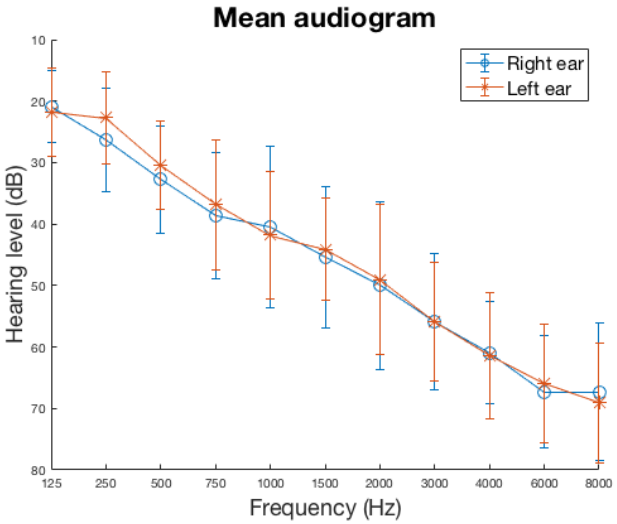


Figure 3.1 The average audiogram for the test subjects. The error bars mark one standard deviation. The hearing loss is in average the same for both ears and stronger for high frequencies, which is common for hearing loss.

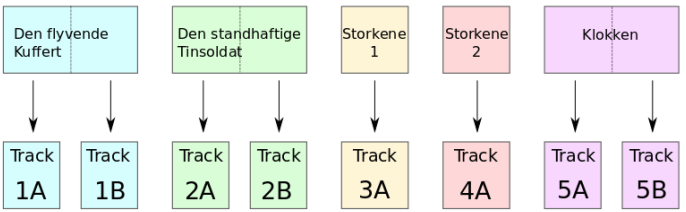


Figure 3.2 An illustration of how some of the stories have been split into two different tracks, and which track corresponds to which story. Each track then corresponds to one EEG recording per test subject.

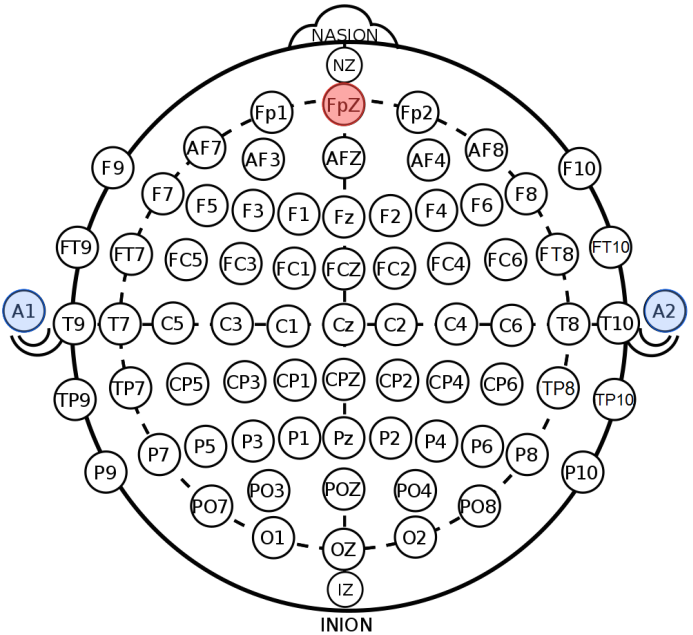


Figure 3.3 An illustration of the electrodes. The red circle marks the main electrode at Fpz, and the blue circles mark the mastoid electrodes at A1 and A2. Modified version of image by [Oxley, 2020].

was recorded using three sintered Ag-AgCl active electrodes. Two of the electrodes were placed on the left and right mastoid respectively, the third was placed on the high forehead on a location called FPz. The placement can be seen in Figure 3.3.

The audio used during the experiment were wav files with a sampling frequency of 44.1 kHz. Since the volume of the sound was noticeably different in the different stories, the files were normalized to a common RMS value. The calibration tone was a 1 kHz pure tone. Besides being played in the loud speaker, the sound was also connected to the EEG system. This results in a second version of the audio signal, called the stimtrack. Because it is recorded by the EEG system, it has the same sampling rate as the EEG signal, which is 16 384 Hz, and was filtered with the same aliasing filter as the EEG prior to digitisation (a low pass filter with a fifth order cascaded integrator-comb (CIC) filter response with a -3 dB point at 1/5th of the sample rate). The recording systems in total causes an 8 ms delay to the recorded signals. The EEG signal has also been processed in the same way as the stimtrack.

4

Methods

4.1 Overview

The work in this thesis mainly consists of two parts. The first is preprocessing of the data, which was done in python. After that, the processed signals were written to binary files that were imported into MATLAB were the estimation of the impulse response was done. A block diagram of what has happened to the signals can be seen in Figure 4.1. This Figure also takes into account the data collection that was described in detail in chapter 3.

This chapter will start by describing how the audio signal was aligned with the EEG signal, which involves finding the onsets of the signals. After that the rest of the preprocessing will be described, including how outliers in EEG signal were handled. After that, the modeling will be described as well as the method that will be used to evaluate the results, the bootstrapping method. Here, there has been a lot of focus on how the algorithms can be tested to ensure that they perform as expected.

Once it has been established that the algorithm perform as expected on simulated data, they were used to find the cortical response. Here some more experiments were performed to test how different methods of preprocessing affected the results.

Finally it is described how the algorithm was applied to find the subcortical re-

Figure 4.1 A block diagram that shows what has been done with the EEG signals and what they have been affected by.

sponse.

4.2 Signal alignment

As mentioned in chapter 3, there exist two versions of the audio signal. One is the original wav file that was played during the experiment. This signal is called audio through out the thesis. The other version of the audio was recorded by the EEG signal and is called the stimtrack. During the experiment, the EEG recording was started before the audio started to play. This means that the audio signal is not aligned in time with the EEG signal and the stimtrack.

The alignment was done in two steps. The first step was estimating the onset of the audio and stimtrack signals. This was done by estimating the derivative in the beginning of the signals, before any sound starts, and marking the first big change in the derivative as the onset of the signal. In this case, big meant a derivative that was more than 10 standard deviations from the mean absolute derivative in the silent part of the signal.

In the second step the shift between the the stimtrack and audio signal was computed by calculating the cross correlation. Since the two signals show the same underlying audio, the time at which the maximal correlation occurs corresponds to the shift between the signals. The shift was computed using code from [Fridman, 2015].

For the above mentioned alignment procedure to be possible, the signals have to be sampled with same rate. As mentioned in chapter 3, however, they are not. According to the documentation, the audio signal is sampled with a sampling frequency of 44,1 kHz while the stimtrack is sampled with 16384 Hz. This means that the signals have to be resampled before doing the alignment. The resampling can, however, be complicated if the recording has been done with systems running on different clocks. If the clocks have drifted in sampling frequency since they were calibrated, this can lead to problems when trying to resample the signals since their sampling frequencies can not be trusted.

Such a drift in the internal system clock can be noticed by computing the shift between the two signals in smaller time windows. If the computed shift is different for windows in the beginning and the end of the signal, then the resampling has not worked properly and the drift has to be estimated. One way of doing this is by assuming that the drift has happened in the clock of the EEG signal. In reality, the drift might have happened in the clock in the audio system, or in both clocks, the assumption is only made to make the estimations easier and when the clock drift is small, as is the case here, later analysis will not rely on the truth of this assumption..

The drift was estimated by plotting the computed shift as a function of time for all tracks and all subjects. The shift was computed in three-seconds-windows, and the resulting shifts were plotted against the start of the time window. The drift of the EEG clock can be found by estimating the slope in the plots. To reduce the influence of outliers, when estimating the slope, the linear model fitting was performed twice, once including all the data points and then a second time that only included the data point that where ± 50 samples from the line of the first model fit. If no data points were found to be used for the second model fitting, the track was excluded for the estimation of the drift. The track was also excluded from the estimation if the threshold estimation could not be performed.

After the alignment was done, the Pearson correlation coefficient between the stim-track and audio signals was calculated to ensure that the alignment had worked as intended. If the correlation between the signals was still lower than 0.7, the recording for that given track and subject was excluded from the modeling. This was the case for 3 out of the 88 tracks.

4.3 Preprocessing the EEG signals

Firstly the EEG signals were referenced by taking the difference between the FPz electrode and the left and right mastoid electrode respectively. This results in two EEG signals, also called EEG channels, and are called FCZ1 and FCZ2. FCZ1 is the difference between FPz and the left mastoid, and FCZ2 to the right mastoid.

The preprocessing was done one recording at a time. This meant that the signals were about 6 minutes long. In order to make the recordings more standardised, they were shortend to be exactly 5 minutes, except for recordings associated with audio track 4A, which were shortend to 4 minutes.

The signals were then filtered, first with a causal first order Butterworth high-pass filter with a bandwidth of 1 Hz. This was done in order to eliminate any drift in the signal.

The signals can be quite far from zero in the initial samples, causing a transient when filtering. Usually this can be avoided by giving the filtering function initial conditions, however, this was not successful in this thesis and the transient remained even when initial conditions were taken into account. Because of this, the transient was removed by removing the first two seconds of each recording.

Since the experiment was conducted in Denmark, the power-line interference is expected to be at 50 Hz and multiples thereof. The interference can be seen in the spectrum which is plotted in Figure 4.2. The EEG signals were thus notch filtered

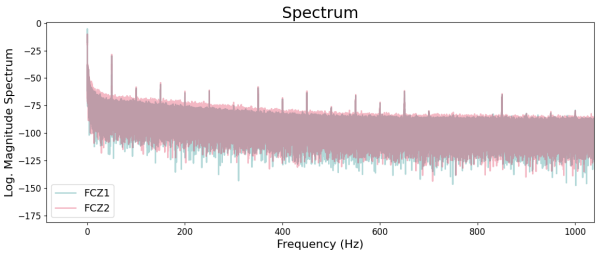


Figure 4.2 An example of what the spectrum of the EEG signals can look like. There are clear peaks visible in 50 Hz intervals which are most likely caused by power line interference.

from 50 to 400 Hz in steps of 50 Hz. The multiples higher than 400 Hz were assumed to be much too high to affect the modeling since the aim is to look for responses with a much lower frequency content. The notch filters were grouped together to two filters with notches at four frequencies each. The reason for using two compound filters, instead of just one, was that numerical difficulties occurred when too many notch filters were put in series. The notch filters were causal IIR filters width a bandwidth of 5 Hz.

After the filtering, outliers were found in the EEG signal. How these were handled is described in greater detail below.

Finally, just before the modeling, the EEG signal was normalized. The reason this was not done earlier was that the preprocessing was done individually for each recording, while the modeling uses all the recordings from one subject. If the normalization was to be done before the recordings had been put together to one long output signal, different normalisations would have been performed on different parts of the signal. The normalization was a z-score normalization, meaning that the mean of the resulting signal is zero and the standard deviation is one.

Outlier removal

The EEG data contains outliers caused by eye blinks and eye movement, among other things. As mentioned in section 2.1, there are multiple ways of handling outliers. In this thesis two different methods were tried and compared. One of the methods is to mark the most extreme values as outliers, and then putting it and the values around it to zero, as was done in [Maddox and Lee, 2018]. In this thesis, the window in which the values were replaced by zeros was 0.5 s wide and centered around the outlier. The other method is to simply mark these values as outliers, but not setting them to zero. In the modeling step the signals will be split into short windows, and windows containing samples marked as outliers are excluded from the modeling.

What percentage of samples that are outliers mainly depends on how much the test

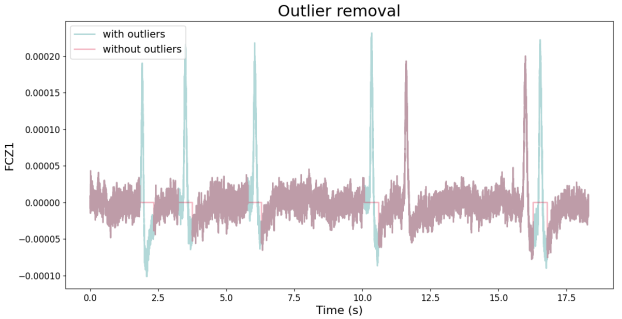


Figure 4.3 The first few seconds of an EEG signal before and after the outliers have been removed. It can be seen that the removal did not manage to remove all artifacts as there is two left, one at about 12 s and one at about 16 s. The threshold for removing therefore has to be altered to remove a higher percentage of the signal.

subject has blinked or moved their eyes. Since this is not channel dependent, the same threshold can mostly be used for both EEG channels. To determine a suitable threshold, the recordings were examined one by one by plotting the EEG signals before and after the removal. An example of such a plot can be seen in Figure 4.3. If the given threshold seemed to have removed all the artifacts without removing too much else, the used threshold was chosen. If too much or little was removed, the threshold was altered accordingly and the removal was redone. Looking at Figure 4.3, it is obvious that several of the artifact were not removed, the threshold had thus to be adjusted to a higher percentage.

4.4 Preprocessing the audio signals

The first step in the preprocessing of the audio signal is aligning it to the EEG signal. How this was done has already been described in detail in section 4.2. After that, the audio signal and the stimtrack signal were processed in the same ways. Firstly, two seconds in the beginning of each track had to be removed in order to keep the signals aligned after the filter transient in the EEG signal has been removed.

When the outliers in the EEG were handled by setting them to zero, the stimtrack and the audio signals were also processed in the outlier removal step. In order to match the regions of zeros in the EEG, the audio signals were also set to zero at the same samples.

As mentioned before, rectifying the signals is thought to help with the modeling since the auditory pathway is indifferent to the sign of the incoming audio [Maddox and Lee, 2018]. To investigate whether this is necessary, it was tried to model both

with and without rectifying the signal. The method for rectifying the input signals was full-wave rectification. This means to invert the sign of all the negative samples, leaving a signal with only positive values. The advantage of using a full-wave rectified signal as oppose to a half-wave is that only one impulse response has to be estimated. Previously, two were estimated, one with the positive values as input, and one with the inverted negative values and the mean of these two estimations was regarded at the estimation of the impulse response [Maddox and Lee, 2018]. Another advantage is that there are less artificial zeros in the input that do not correspond to zeros in the output, this might make estimation easier.

Similarly to the EEG signal, the input signals were also normalized just before the modeling took place.

4.5 Impulse estimation

As mentioned before, the estimation was done in MATLAB. The estimation was done according to the theory presented in section 2.6. The estimation was done with windowing using a Hann window. This meant dividing the signals into smaller segments, called windows, applying a Hann window to those segments before doing the estimation of the frequency response. The mean frequency for all windows is then estimated and this mean is used to estimate the impulse response. One important reason for dividing the signals in segments is that the data can be assumed to be stationary for each segment. While the signals in their entirety are not stationary, short segments are almost stationary.

As mentioned there, the parameter ε decides which frequencies that are used when estimating the frequency response. Because the spectrogram of the real signal is very uneven it is hard to choose an ε so that only frequencies of interest are included, therefore, ε was chosen ad hoc to be 200. This roughly translated to a threshold of a couple of hundred Hz for the window length 1.3 s, which was used when estimating.

Testing the algorithm

Before the above described algorithm was used on the real data to estimate subcortical impulse responses, a number of tests were performed. These tests' main purpose was to ensure that the algorithm worked as intended. In addition, some tests were performed to examine the effects of the estimation parameters.

The first step is to ensure that the algorithm can find a known impulse signal. For this an input signal was convolved with a known impulse response. To make the problem as similar to the original one as possible, the input signal was the preprocessed audio signal and the known impulse response looked similar to the traditional ABR

in Figure 1.1. After the convolution, white noise with a small variance was added. This gives the final signal high SNR compared to a traditional ABR measurment.

Once it had been established that the algorithm can find a known impulse response, different values ε were tried. The effect of the window length was also examined.

4.6 The bootstrapping method

An important question is that of determining whether a found impulse response is significant. This problem is very similar to that of determining whether or not a response is present when the traditional ABR is measured. One method of doing this is with bootstrapping [Lv et al., 2007]. Bootstrapping methods relay on a small amount of data that is used to create many different data sets. In this thesis, bootstrapping is used to create a null hypothesis of what a parameter value should be if no response is present. This null hypothesis can then be tested with the original data to determine if it is significantly different from when there is no response. If it is significantly different, the null hypothesis, that there is no response, can be rejected. This method of testing a null hypothesis works just like described in section 2.5, just that the distribution is unknown and therefore has to be estimated by simulation.

Similarly to [Aljarboa et al., 2022], the bootstrap data sets are created by misaligning the signals a random amount, but no shorter than 15.6 s, in accordance with [Aljarboa et al., 2022]. By misaligning the signals enough, there should not be any influence of the audio on the brain activity. This means that the newly created data is data without an underlying impulse response to find. It should be mentioned here that [Aljarboa et al., 2022] applied this method on determining the significance of cortical responses, not subcortical ones. The method should still work, however.

The misaligning was done by moving a random amount of samples of the input signal from the beginning of the signal to the end. This meant that the algorithm tries to find correlation between the brain activity and sound that will be heard in the future, meaning that (causal parts of) the impulse responses that are found should non-significant. Each impulse response is then used to estimate the output signal. This estimated output signal was then compared with the real output signal by computing the Pearson correlation coefficient between them, this value is called estimation accuracy. In this thesis, this procedure is repeated 100 times to get an estimate of the distribution of the estimation accuracy when there is no correlation between the input and output signals. This distribution is then used to obtain a threshold to test the null hypothesis of no response in the original aligned data. The p-value of the confidence interval was chosen to be 0.05, meaning that the threshold was chosen such that the estimation accuracy had to be greater than 95% of the estimation accuracy that were computed for the misaligned data.

Testing the bootstrapping method

The method was tested with white noise as input and output to ensure that the false-positive rate was consistent with the chosen p-value. This was done by creating 100 input-output data sets of random noise and running each of them through the bootstrapping algorithm to obtain a threshold for each data set, and then comparing the estimation accuracy obtained for the aligned data with the threshold. Since the data is just noise, anytime a response is deemed significant is a false-positive.

The bootstrapping algorithm was also tested when there was a response to find. To do this, a known impulse response was convolved with a white noise input signal to obtain an output signal without noise. On top of this output signal different levels of white noise was added to get a noise output signal that was used when estimating the impulse response. This gives insight into how low the SNR can be for the estimation to work. For each SNR, ten simulations were done. The results of these tests can be seen in Section 5.2 and they showed that the algorithm worked properly unless there was a lot of noise added.

4.7 Finding the cortical response

While the aim of this thesis is to find the subcortical response, the algorithms were also used to look for the cortical response. This served as a final test for the algorithm using the real data. Since the cortical response has a larger amplitude than the subcortical one, the SNR in this problem is higher, which makes the problem easier.

To adjust the algorithm to find the cortical response, the signals were down sampled to a lower sampling rate. This could be done since the cortical response has a lower frequency content and a lower sampling rate leads to faster estimation. This is especially valuable since the impulse response that has to be estimated, spans more time lags than the subcortical impulse response. As mentioned in section 2.3, the long latency responses last up until 200 ms after the stimulus. To ensure that everything worked as intended, the impulse response was estimated from -100 ms to 1 s. This way, it is not only possible to examine if there is a response when it is expected, but also check that there is no response when it is not expected.

The bootstrapping method was then used to determine whether the found cortical response were significant or not. The estimated impulse responses was also visually compared with traditional cortical activity in the AEP.

The modeling was done with some different preprocessing methods to compare how they affected the data. There was one comparison between the two different ways EEG outliers were handled in this thesis, as well as an experiment to determine whether the input signals should be rectified or not. First the experiment with the

inputs was done and the outliers were handled by ignoring windows containing outliers. Then rectified inputs were used to compare the methods of handling outliers. The results from this can be seen in Section 5.3.

Once a suitable preprocessing method had been found it was applied to the signal and the whole experiment was then repeated but with reversed input data. Since the signals are a lot longer than the assumed impulse response, there should not be any correlation between the reversed input signals and EEG signal. This means that the bootstrapping algorithm should not deem the result to be significant and the estimated responses should look like noise. The results from this can also be seen in Section 5.3.

4.8 Finding the subcortical response

Once it has been established that the algorithm works as intended and what preprocessing worked best, the final step was to try it on the real data to estimate the subcortical impulse response. The procedure was essentially the same as it was when the cortical response was estimated. Only this time, the signals were not down sampled and the response is estimated for time lags -20 ms to 50 ms. Another difference to the the cortical estimations, the found impulse responses were low-pass filtered to obtain a smoother impulse response. The responses were filtered with four with first order Butterworth, first one with a bandwidth of 2000 Hz, followed by ones with bandwidths of 1000, 500 and 100 Hz.

Again the estimation accuracy of the estimated impulse response model was compared to the threshold computed by the bootstrapping algorithm to determine the significance of the response. The estimated impulse responses were then visually compared to the traditional ABR. Similarly to when the cortical responses were estimated, the experiment was then repeated using reversed input data.

5

Results discussion

5.1 Preprocessing of the signals

Sampling rate drift

An example showing the shift between the audio and stimtrack after resampling can be seen in Figure 5.1. It shows how the shift is different between the start of the signals and the end of the signals, indicating that the resampling has not lead to signals with the same sampling frequency.

The drift in sampling frequency was estimated to be 0.6 Hz with a standard deviation of 0.05 Hz. The estimation was done using 62 out of the 88 tracks.

Taking the drift into account when resampling, and again estimating the shift between the signals at different times in the signal, gives that there is no change in shift along the signal. This is illustrated in Figure 5.2.

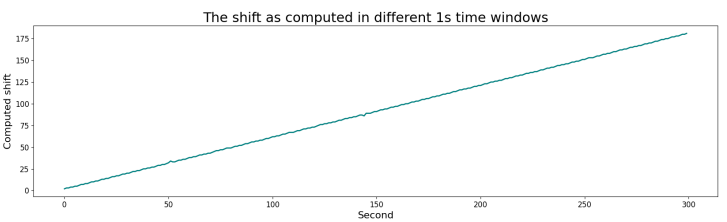


Figure 5.1 The shift as computed in different time windows along the stimtrack and audio signal, when the documented sampling frequencies are assumed to be correct. The fact that it is a linear trend with a non-zero slope shows that the two signals do not have the same sampling-rate despite the resampling. The value of the slope corresponds to the drift between the two sampling rates. The specific recording used for this plot was chosen so that the slope is clearly visible and is not the same recording as in Figure 5.2.

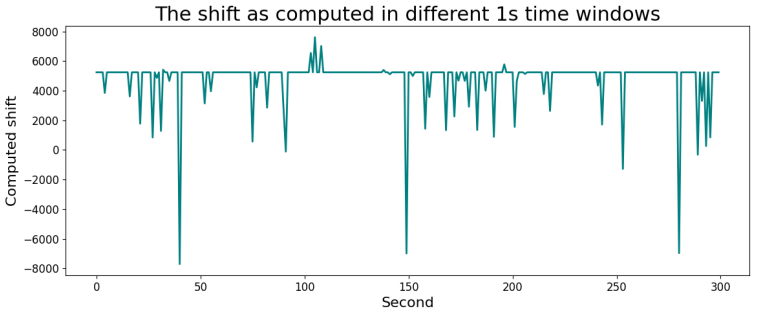
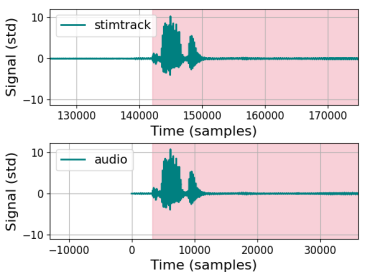
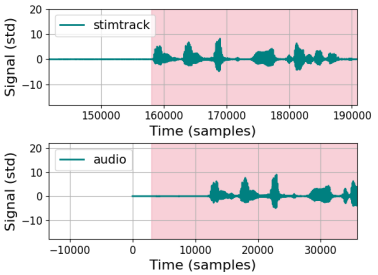


Figure 5.2 The shift as computed in different time windows along the stimtrack and audio signal, when the estimated drift has been taken into account. The fact that there is no trend in the shift suggests that the signals now are sampled with the same frequency. There are a lot of outliers though. The specific recording used for this plot was chosen so that the slope is clearly visible and to show the type of outliers that can be found when computing the shift. It is not the same recording as in Figure 5.1, which did not show any outliers.



(a) Test person TP11 track 5A. The alignment has worked well.



(b) Test person TP11 track 2A. The alignment has not worked well.

Figure 5.3 Two examples of results of the first alignment that estimates the onset of a signal by looking for the first big change in the signal. The red area marks where the signal is estimated to begin. It is evident that the alignment works well for some signals (left), but not others (right).

Signal alignments

The first alignment method, that estimated the onset of the signal, worked well for some signals, see e.g. Figure 5.3a, but less well in others, see Figure 5.3b.

The second alignment method, that computes the shift between the stimtrack and audio, leads to a better alignment for the trials that were misaligned after the threshold estimation, compare Figure 5.4 with Figure 5.3b.

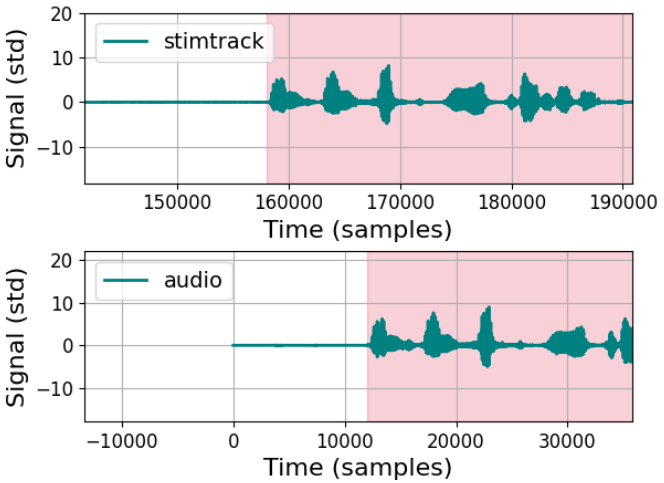


Figure 5.4 The signals of Test person TP11 track 2A, after the alignment has been adjusted by computing the shift between the signals using correlation. The red area marks where the signal is estimated to begin. The alignment has worked well.

The mean Pearson correlation coefficient between the stimtrack and the audio signal after the two alignment methods was 0.986 and the standard deviation was 0.021. This suggests that the alignment worked well. These values were computed for 85 out of the 88 tracks.

5.2 Simulated data

Testing the impulse estimation

The results from simulated data with some different parameter values can be seen in Figure 5.5 and Figure 5.6.

The estimated impulses in Figure 5.5 and Figure 5.6 look similar to the real underlying impulse response. This suggests that the algorithm works well for finding a known impulse response. It is clear, however, that the parameter ε affects the ability to find the correct impulse response while the window length does not seem to affect the estimation.

The greater ε, the further from the real impulse response the estimation is, in Figure 5.5. This is typical for regularization parameters were you usually accept a larger bias to the estimate in order to get a lower estimation error overall. This is only effective in noisy signals and in this simulated case there is no advantage to having

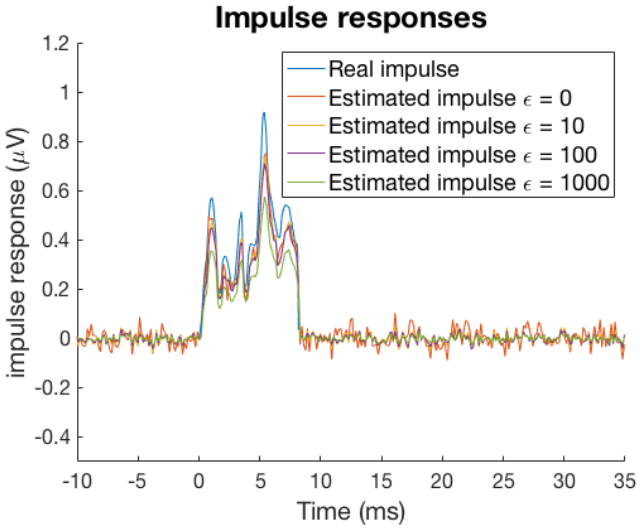


Figure 5.5 The effect ϵ has on the estimation. The lower the epsilon, the closer to the real solution the estimation is.

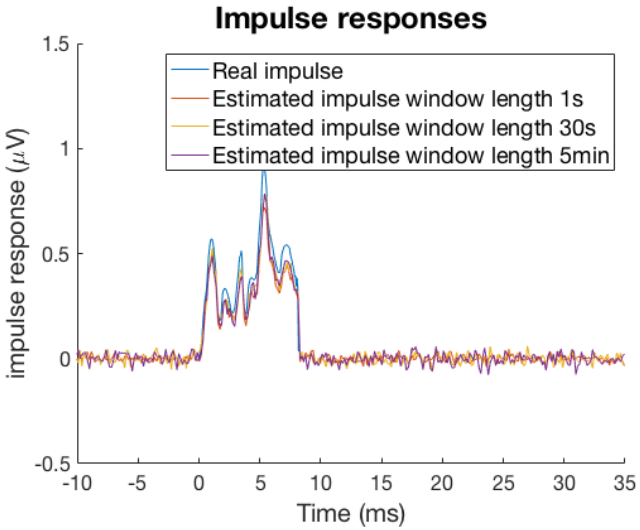


Figure 5.6 The effect window length has on the estimation. There is no clear difference between the different window lengths.

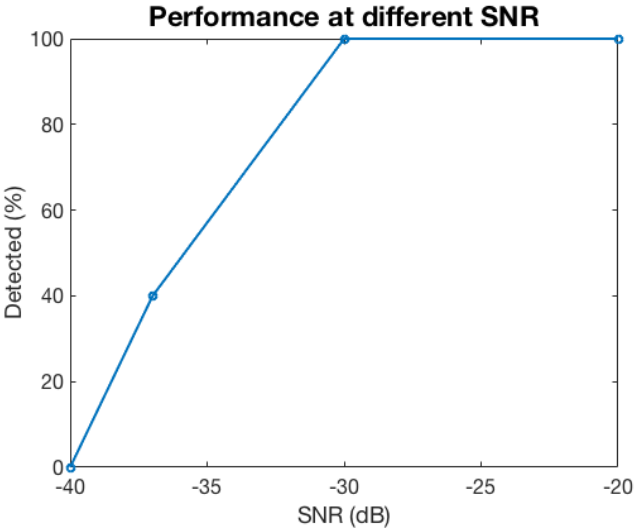


Figure 5.7 The percentage of simulations that were deemed to be significant by the bootstrapping level when there is an impulse response to find among the noise. In other words, the true-positive rate at different SNRs. For each SNR, ten simulations were done

a large ε.

Testing the bootstrapping method

When running the bootstrapping algorithm with white noise as input and output, 5 out of 100 simulations resulted in that the bootstrapping method determined the result to be significant. This is in accordance with the p-value, which was chosen to be 0.05. This is one sign that the algorithm is working as expected.

When the output was obtained by convolving the input with an impulse response, and then adding noise, the algorithm found impulse responses that were deemed significant if the noise level was not to high. Results for the different noise levels can be seen in Figure 5.7

Interestingly, if not surprisingly, the results from Figure 5.7 suggests that the method struggles with finding existing significant results if the SNR is too low. This is very relevant since the ABR is notorious for having low SNR.

Table 5.1 The number of significant impulse responses depending on whether the input signal was rectified or not. For each category there are four impulses meaning that the sum for each method is at most 44. More significant impulse responses are found if the input signals have been rectified using full wave rectification.

Test person	Significant impulse responses with rectified inputs	Significant impulse responses with non-rectified inputs
TP01	4	0
TP02	1	0
TP03	4	3
TP04	3	1
TP05	4	0
TP06	4	1
TP07	0	1
TP08	2	1
TP09	0	0
TP10	2	1
TP11	4	0
Total	28	8

5.3 Cortical response

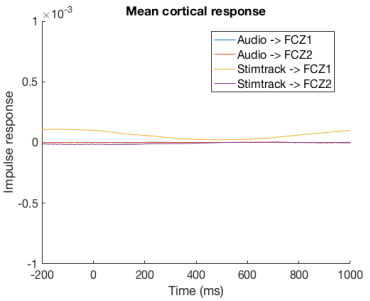
Rectifying the input signal

The number of significant responses for the rectified and non rectified signals can be seen in Table 5.1. It is evident that more of the responses are deemed significant when the input signals are rectified before modeling.

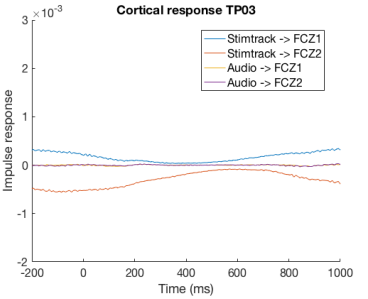
Figure 5.8 shows what the impulse responses look like depending on whether the input signals have been rectified or not. Figures 5.8a and 5.8c show the average response for all responses that were deemed significant for the non rectified and rectified case respectively. Figures 5.8b and 5.8d give an example of what the impulse responses look like for one of the participants.

Figure 5.8 shows that when the inputs were not rectified, there were no features at all to see in the impulses. They all looked more or less flat. When the full-wave rectification was used, however, some features were visible.

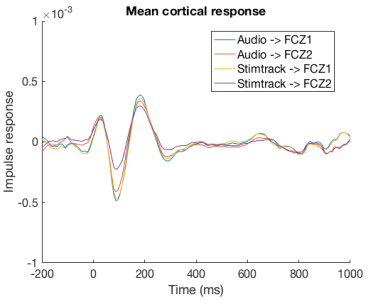
The features in Figure 5.8c are especially interesting. It has three clear peaks; a positive around 25 ms, a negative around 87 ms, and another positive around 190 ms. This can be compared with the peaks in the traditional long-latency response P50, N100 and, P160 as seen in Figure 2.3. While the pattern is pretty much the same, the two first peaks are earlier than expected and the last one later. The earlier peaks are especially surprising considering that there should be about a 15 ms delay



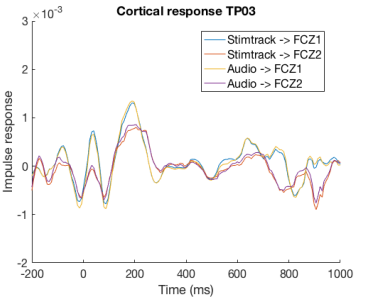
(a) The mean impulse responses for all subjects when the inputs were not rectified.



(b) Non rectified inputs for test person TP03. Note that the impulse for system audio to FCZ2 was not deemed significant.



(c) The mean impulse responses for all subjects when the inputs were rectified using full wave rectification.



(d) Rectified inputs for test person TP03. All impulse responses were deemed significant

Figure 5.8 A visual comparison of what the impulse responses look like depending on whether or not the input signals have been rectified. While it is hard to see anything in the impulse responses when the signals have not been rectified, the impulse responses for the rectified input share features with the traditional cortex response, namely the waves resembling P50, N100 and, P160.

due to the time the sound has to travel from the loudspeaker, as well as the delay in the EEG recording system which was reportedly 8 ms.

The offset in the peaks could be explained by filtering, although, the filters that were examined did not show any clear delays. Furthermore would the filter have to affect the different peaks differently, which is possible if the phase response is nonlinear.

It should be mentioned that the peaks in the cortical impulse response do not necessarily correspond to the peaks in the long-latency AEP. It is possible that they stem from different processes in the brain. This could also be one explanation for why the peaks do not occur at the same lags.

Over all rectifying the signals seems to be an important step. This is in line with what was stated by [Maddox and Lee, 2018]. Presumably, since the auditory pathway is indifferent to the sign of the incoming sound, the auditory pathway would react the same way to the full wave rectified signal as to the original signal. Since the impulse response is a linear model, it can not capture this nonlinear feature and therefore the modeling does not work unless the nonlinearity is added before the modeling.

Outlier removal

How much that was removed of the data due to outliers varied a lot from subject to subject and on the method for handling the outliers. Table 5.2 shows how much was removed for each subject.

Using the entries in Table 5.2 it can be calculated that the mean data that was removed is 42% when the outliers are set to zero and, 66% when windows containing outliers are ignored. That is to say, more data is lost when entire windows are ignored. This, of course, makes sense since what is removed are the outliers alongside all the other data in the same window.

How the method of handling outliers affects the number of significant impulse responses can be seen in Table 5.3. When the outliers are set to zero, all the found impulse responses are deemed to be significant while only slightly more than half are deemed significant if the windows containing outliers are ignored.

Figure 5.9 shows what the impulse responses look like depending on which method was used for handling the outliers. Figures 5.9a and 5.9c show the average response for all responses that were deemed significant for the two methods respectively. Figures 5.9b and 5.9d gives an example of what the impulse responses look like for one of the participants, TP05.

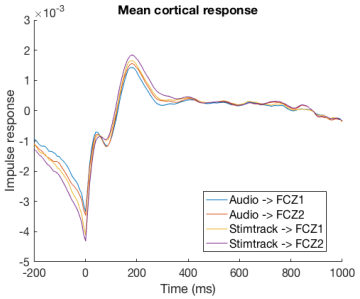
When just looking at how many significant impulse responses that were found, ze-

Table 5.2 How much of the signals that remains for each test subject, after the artifacts have been removed. For some subjects, e.g. TP01, a lot of the signal remains, while others, e.g. TP07, barely have anything left after the artifacts have been removed.

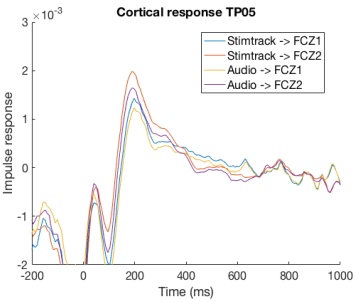
Test person	Ignoring windows containing outliers		Setting outliers to zero	
	Remaining in FCZ1 (%)	Remaining in FCZ2 (%)	Remaining in FCZ1 (%)	Remaining in FCZ2 (%)
TP01	83	81	90	88
TP02	25	26	67	66
TP03	92	92	93	94
TP04	63	60	71	69
TP05	16	16	52	51
TP06	20	22	45	45
TP07	1	1	24	27
TP08	1	2	30	38
TP09	2	2	33	34
TP10	56	57	79	81
TP11	18	12	50	57

Table 5.3 How the method of handling outliers affects the number of impulse responses that are deemed significant. For each category there are four impulses meaning that the sum for each method is at most 44. It is evident that setting the outliers to zero results in more impulse responses to be deemed significant. The left column in this Table shows the same results as the left collumn in Table 5.1

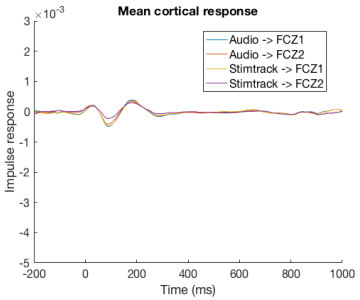
Test person	Ignoring windows containing outliers	Setting outliers to zero
TP01	4	4
TP02	1	4
TP03	4	4
TP04	3	4
TP05	4	4
TP06	4	4
TP07	0	4
TP08	2	4
TP09	0	4
TP10	2	4
TP11	4	4
Sum	28	44



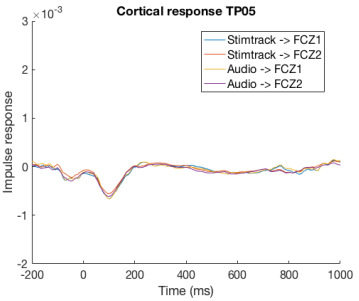
(a) The mean impulse responses for all subjects when the response was deemed significant and the outliers were set to zero.



(b) Test person TP05 when the outliers are set to zero.



(c) The mean impulse responses for all subjects when the response was deemed significant and the windows containing outliers have been ignored.



(d) Test person TP05 when the outliers are windows containing outliers have been ignored.

Figure 5.9 A visual comparison of what the impulse responses look like depending on what method is used for handling the outliers. When the outliers are set to zero the amplitude of the response is larger and it contains a large negative peak at lag zero. When all windows containing outliers are ignored the peaks are weak but could correspond to features in the traditional cortical response, namely the waves P50, N100 and, P160. The waves are however, not exactly were they are expected.

roring the outliers might seem like the favourable method for treating them. When looking at the found impulses, however, it is clear that the method has problems, meaning that the bootstrapping method can not be trusted blindly.

The mean impulse response for zeroed outliers in Figure 5.9a has a very large negative peak at zero that continues to negative time lags, making the estimated impulse response acausal. Since the EEG signals should not be affected by sounds that will be heard in the future, this acausal impulse response is faulty.

To investigate the non-causal peak in more detail, the responses for the rest of the test subjects can be seen in Figure 5.11.

When comparing the plots in Figure 5.11 with Table 5.2 there seems to be a trend that the subjects where a lot of data was removed are the subjects with large negative peaks at lag zero. Add to that the fact that the large peak at zero disappears if the windows containing outliers are ignored. This suggests that the peak is a result of the outlier handling.

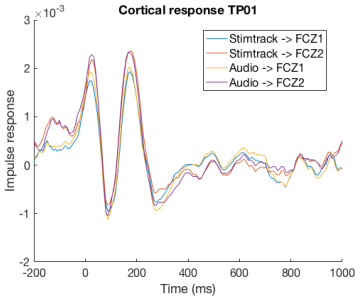
It is plausible that this is an effect of that not only the EEG signal outliers were set to zero, but that also the input was set to zero for the same lags. This was done in order to not have inputs that resulted in a zero output since this was thought to confuse the algorithm. The problem might stem from that setting the same samples to zero in both the input and the output suggests to the model that the input has an immediate effect on the output, which would result in a correlation at lag zero. Because of this, it would be interesting to investigate how well it works to put the outliers to zero but without putting the input to zero.

A lot of data was removed in the outlier removal, in some instances so much that it is questionable whether there is any use in keeping the data from that test subject since there is barely anything left. This definitely affects the results and it is quite possible that setting the outliers to zero is acceptable if there are less outliers. Seeing that this is not the case here, it was chosen to handle the outlier by ignoring windows containing them. This way a lot more of the data is removed but there are no clear artifacts due to the outlier removal in the estimated impulse response.

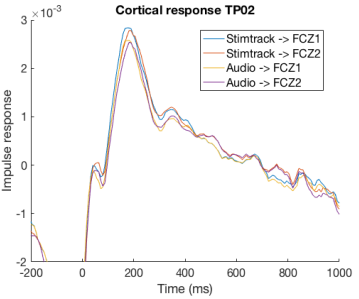
Reversing the data

These final cortical response estimations were done using rectified input signals and by ignoring windows containing outliers. This was because these preprocessing methods were deemed to be the best based on the results so far.

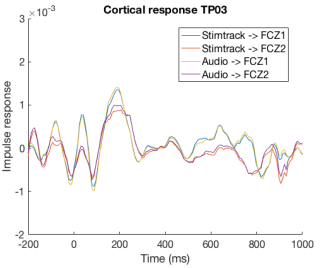
The thresholds that were computed by the bootstrapping method were different from subject to subject. The mean threshold was an estimation accuracy of 2.4% across all systems with the original input signal. This is a very low threshold. For the



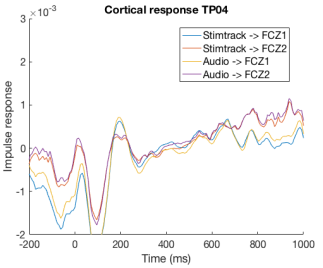
(a) TP01



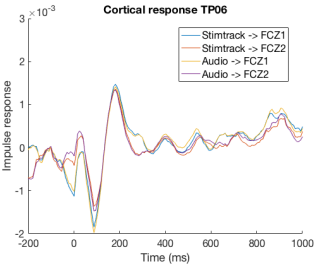
(b) TP02



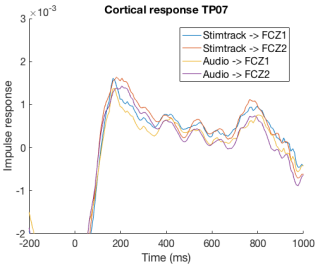
(c) TP03



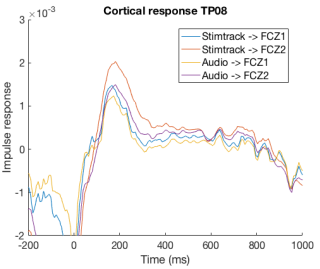
(d) TP04



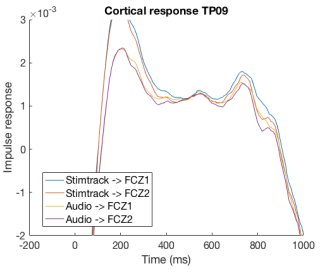
(e) TP06



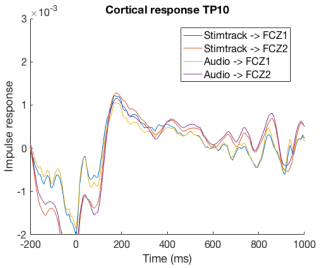
(f) TP07



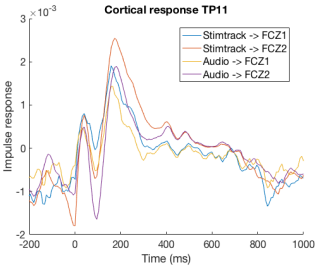
(g) TP08



(h) TP09



(a) TP10



(b) TP11

Figure 5.11a The estimates responses when outliers are set to zero for the different test subjects. Some of the figures show large negative peaks, e.g. TP02. When comparing these results with Table 5.2 there seems to be the trend that the subjects where a lot of data was removed are the subjects with large negative peaks.

reversed input signals the mean threshold was 2.3%, almost the same as for the real input signal. Furthermore the estimation accuracy was never more than 9%. This is not completely unreasonable since there is a lot going on in the brain that has nothing to do with the sound you hear.

It is also worth noting that the test subject with less removed data to a greater extent seem to have more significant impulse responses. For instance were four significant responses found for TP01 with 83% and 81% remaining in the two EEG signals respectively, while none were found for TP09 with 2% and 2% remaining. This is no hard rule, however, since there were e.g. one significant response for TP02 and four for TP05 despite TP02 had slightly less data removed.

In total 29 of the impulse responses were deemed significant, when the data was not reversed, and 1 when the data was reversed.

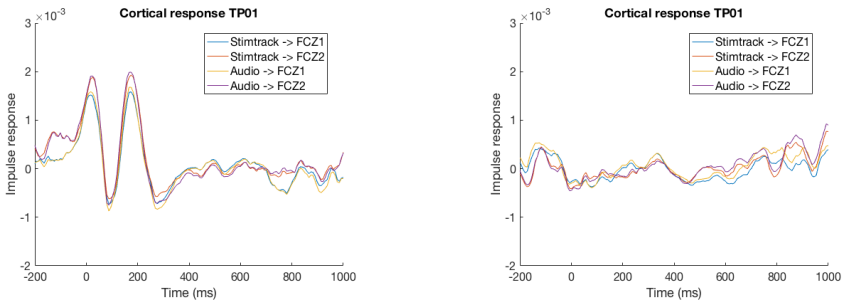
For an example of how the threshold were for one subject, see Table 7.1. There you can also see the estimation accuracy values for the original, correctly aligned data, which is written in a bold font whenever it is larger than its corresponding threshold, meaning that the estimated impulse response is deemed significant. Similar tables, for the other test subjects, can be found in the Appendix A. The impulse responses, that correspond to Table 7.1, can be seen in Figure 5.12.

From Figure 5.12, it is evident that the impulse responses for the reversed input signal are smaller than those for the original input signals for positive lags. All the impulse responses, however have about the same amplitude for negative lags.

The estimated impulse responses for all subjects can be seen in Figure 5.14.

Table 5.4 The results from the bootstrapping method when applied to the data from TP0899 to obtain the cortical response. All the responses for the real data are deemed to be significant, while the responses for the reversed input signals are not. The notation ^{r} indicates that the signal has been reversed and the bold numbers indicate that the estimation accuracy is large enough be be deemed significant.

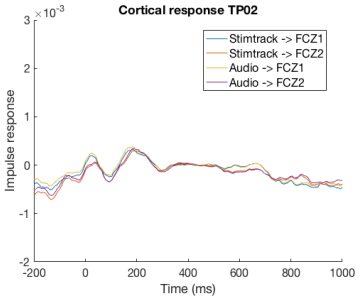
System	Threshold	estimation accuracy
Stimtrack $\rightarrow$ FCZ1	0.0173	**0.0718**
Stimtrack $\rightarrow$ FCZ2	0.0118	**0.0769**
Audio $\rightarrow$ FCZ1	0.0156	**0.0797**
Audio $\rightarrow$ FCZ2	0.0143	**0.0864**
Stimtrack^r $\rightarrow$ FCZ1	0.0149	0.0069
Stimtrack^r $\rightarrow$ FCZ2	0.0140	0.0069
Audio^r $\rightarrow$ FCZ1	0.0154	0.0043
Audio^r $\rightarrow$ FCZ2	0.0144	0.0078



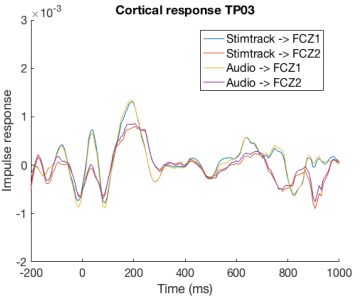
(a) Real input signals

(b) Reversed input signals.

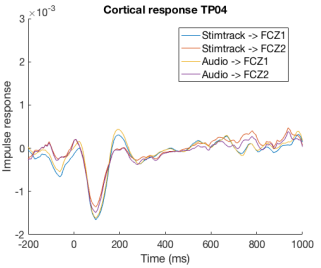
Figure 5.12 An example of the estimated impulse responses for both the original input signals, as well as for the reversed signal. The responses for the original input data have a much larger amplitude for the positive lags, while the negative lags have about the same values as the values of the impulse responses with the reversed input signal. All the systems with the original data were deemed significant by the bootstrapping method, while those for the reversed data, most of the time, were not.



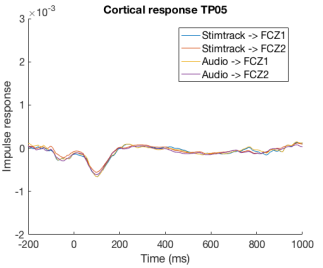
(a) TP01



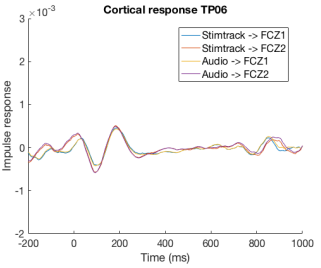
(b) TP02



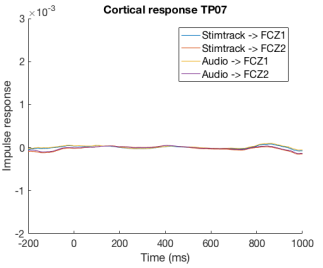
(c) TP03



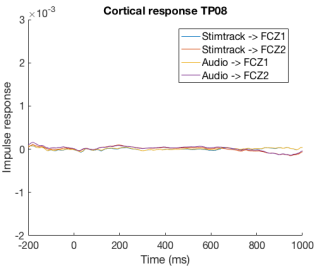
(d) TP04



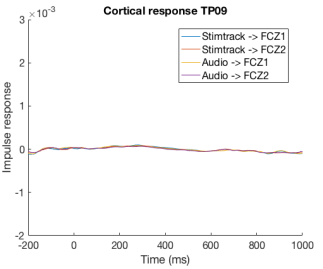
(e) TP06



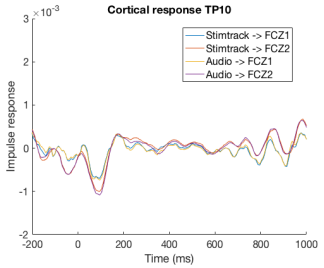
(f) TP07



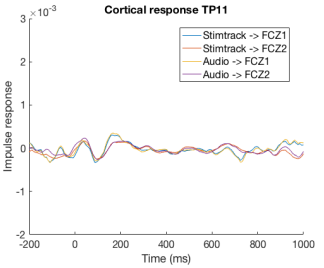
(g) TP08



(h) TP09



(a) TP10



(b) TP11

Figure 5.14a The estimates impulse responses with the final preprocessing method, rectifying the signal and ignoring windows containing outliers. Many of the plots share features with the long-latency AEP but some are more or less featureless. Comparing with Table 5.2 it seems that the flat impulse responses correspond to when a lot of data was removed.

Many of the plots in Figure 5.14 show feature similar to the long-latency AEP, just like the mean response did in Figure 5.8c. Some response, however, have no clear features and look more or less flat. When comparing with Table 5.2 it seems like the flat responses correspond to when a lot of data was removed.

Over all, the results suggest that modeling has found an underlying impulse response and that the algorithm seems to work on real data.

5.4 Subcortical response

The results in this section use the same preprocessing as for the final results when estimating the cortical response, i.e. rectifying the input signals and handling the outliers by excluding windows containing outliers when estimating.

Table 5.5 shows how much of the data that was ignored due to outliers. Just as for the cortical response, a lot of the signals are unused, especially for some participants e.g. TP07 were only 1% of the signal remains.

Similarly to the cortical response, the thresholds that were computed by the bootstrapping method were different from subject to subject. The mean threshold was a estimation accuracy of 1.16% across all systems with the original input signal. For the reversed input signals the mean threshold was 1.15%.

An overview over the results from the bootstrapping method can be seen in table 5.6. It shows how many significant impulse responses were found for the original data, and the reversed data, respectively. For each subject, the maximal number of impulse responses that were estimated was eight, four with the real input, and four

Table 5.5 How much of the data that was ignored in the analysis of the subcortical response due outliers. It is evident that there is barely any data left for some of the participants e.g. TP07. compare with Table 5.2 for the cortical response

Test person	Remaining in FCZ1	Remaining in FCZ2
TP01	84	82
TP02	28	28
TP03	92	93
TP04	64	61
TP05	18	18
TP06	22	23
TP07	1	1
TP08	1	2
TP09	2	2
TP10	60	60
TP11	20	15

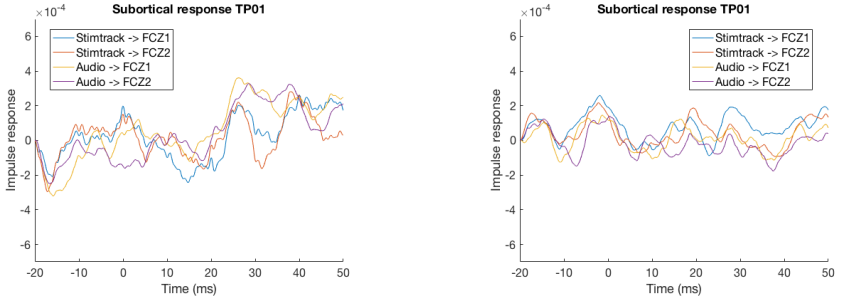
Table 5.6 How many impulse responses that were deemed significant for the reversed an nonreversed inputs signals. For each category there are four impulses meaning that the sum for each method is at most 44. There were very few responses found for the nonreversed data and non for the reversed.

Test subject	Impulse responses in real data	Impulse responses in reversed
TP01	4	0
TP02	0	0
TP03	0	0
TP04	0	0
TP05	0	0
TP06	0	0
TP07	0	0
TP08	1	0
TP09	0	0
TP10	1	0
TP11	0	0

with reversed input.

The results from Table 5.6 show that barely any significant impulse responses were found for the subcortical case.

An example of what a significant impulse response can look like can be seen in Figure 5.15a. The impulse responses for the reversed inputs can be seen in Figure



(a) Real input signals **(b)** Reversed input signals.

Figure 5.15 An example of the estimated impulse responses for both the original input signals, as well as for the reversed signal. None of the impulse responses have any clear features corresponding to peaks in the ABR.

5.15b.

None of these impulse responses in Figure 5.15a have clear features corresponding to peaks in the traditional ABR in Figure 1.1. Again, it should be remembered that there is an expected delay of about 15 ms. This would mean that the strongest peak, wave V, should be located around 20 ms.

Furthermore, there is no great visual difference between the impulse when the input signals are reversed or not and the impulse responses for the different system look more different from each other than they did for the cortical response.

5.5 Stimtrack or audio signal

While the performance is more or less the same, the preprocessing was significantly more involved for the audio signal. Since the stimtrack is recorded by the EEG system it is sampled with the same sampling rate and is started at the same time as the EEG recording. This means that there is no need for aligning the signals and the issue with drifts in the sampling clocks are avoided. This makes the stimtrack a more favourable input signal.

5.6 Future work

A major problem with the analysis in this thesis is the handling of outliers. The data set contains a lot of artifacts due to eye movement, blinks and the electrical activity of the heart and these are handled in very simple ways that remove a lot of the data. It is possibly because of this that the algorithm did not work for finding the subcortical response. Furthermore, the methods are not automated since it requires

someone who determines a suitable threshold for the outlier removal. This is very labour intensive. The main focus of future research should therefore be on finding better methods for handling the outliers. Since the outliers are distinctly different from the rest of the EEG signal it might be possible to identify them and subtract them from from the signal instead of throwing away the data completely. As already mentioned, it would also be good to investigate how well it works to put outliers to zero but without putting the input to zero for the same values.

If a new data set is collected, this could be done with artifacts in mind. For instance could the eye movement be measured with an EOG and the electrical activity of the heart with an ECG. This signal could then be used for artifact cancellation. While this approach might be infeasible when implementing this technology in a hearing aid, it could help in examining if it is possible to estimate hearing thresholds from these impulse responses.

Another alternative, if more data is collected, is to use more EEG channels when during the recording. This would allow for ICA to filter out artifacts due to eye-movements or heart activity.

Future work should also look into different ways of evaluating the results from the estimation. In this thesis the Pearson correlation coefficient between the real output and the estimated output was used in combination with a bootstrapping method. While the method seems to work as intended on simulated data and it is a good check to determine if there is something, there is no guarantee that the something that is found is of interests. This was obvious when different methods for handling outliers were compared and there were peaks in the impulse response that presumably stemmed from the outlier removal. Despite the peaks not being associated with brain activity the bootstrapping method deemed all responses to be significant. One can therefore not blindly trust the bootstrapping method to verify results.

Another interesting area for future work would be to look more into how the algorithm performs for simulated data with a lot of noise. This could be pared with experiments that estimate the SNR for a traditional ABR experiment. This could give more insight into whether this method is feasible for finding ABR like impulse responses.

6

Conclusion

In conclusion, this thesis explored the feasibility of estimating the subcortical response by estimating the impulse response from the audio to the EEG signal. This was done to investigate a possible method of determining someones hearing threshold that could be implemented in an adaptive hearing aid to tune its features to better fit the user's needs.

Some different methods of preprocessing were investigated. For one, it was determined that it is important to rectify the input signal before estimating the impulse response. This mimics the fact that the auditory pathway reacts the same way to positive and negative stimulus. It was also suggested that replacing EEG outliers with zeros, as well as setting the input to zero for the same samples, results in artifacts in the estimated impulse response. A better approach is therefore to reject signal windows that contain outliers when doing the modeling although more data is lost this way.

While the implemented algorithm worked well on simulated data and managed to find significant responses similar to the cortical response, no results were found that seemed to correspond to the subcortical response. One possible reason for this might be that a great deal of data was removed from the data set because it contained outliers. Future research should therefore focus on finding better ways of handling outliers.

7

Appendix

A Cortical responses

The tables in this appendix show the results from when the bootstrapping method was applied when estimating the cortical response. The notation ^{r} indicates that the signal has been reversed and the bold numbers indicate that the estimation accuracy is large enough be be deemed significant. The definition of estimation accuracy can be found in Section 4.6.

Table 7.1 The bootstrapping test on cortical activity for subject TP01 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack $\rightarrow$ FCZ1	0.0192	**0.0718**
Stimtrack $\rightarrow$ FCZ2	0.0142	**0.0769**
Audio $\rightarrow$ FCZ1	0.0152	**0.0797**
Audio $\rightarrow$ FCZ2	0.0125	**0.0864**
Stimtrack^r $\rightarrow$ FCZ1	0.0153	0.0102
Stimtrack^r $\rightarrow$ FCZ2	0.0113	0.0095
Audio^r $\rightarrow$ FCZ1	0.0156	0.0080
Audio^r $\rightarrow$ FCZ2	0.0146	0.0087

Table 7.2 The bootstrapping test on cortical activity for subject TP02 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0263	0.0133
Stimtrack → FCZ2	0.0239	0.0238
Audio → FCZ1	0.0252	0.0168
Audio → FCZ2	0.0258	**0.0282**
Stimtrack^r → FCZ1	0.0236	0.0038
Stimtrack^r → FCZ2	0.0232	0.0029
Audio^r → FCZ1	0.0181	0.0041
Audio^r → FCZ2	0.0249	0.0053

Table 7.3 The bootstrapping test on cortical activity for subject TP03 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0118	**0.0282**
Stimtrack → FCZ2	0.0111	**0.0290**
Audio → FCZ1	0.0126	**0.0294**
Audio → FCZ2	0.0137	**0.0296**
Stimtrack^r → FCZ1	0.0121	0.0100
Stimtrack^r → FCZ2	0.0102	0.0017
Audio^r → FCZ1	0.0113	0.0105
Audio^r → FCZ2	0.0102	0.0031

Table 7.4 The bootstrapping test on cortical activity for subject TP04 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0270	**0.0441**
Stimtrack → FCZ2	0.0317	0.0196
Audio → FCZ1	0.0193	**0.0464**
Audio → FCZ2	0.0194	**0.0374**
Stimtrack^r → FCZ1	0.0301	0.0297
Stimtrack^r → FCZ2	0.0289	0.0134
Audio^r → FCZ1	0.0224	-2.7712e-04
Audio^r → FCZ2	0.0203	-4.0971e-05

Table 7.5 The bootstrapping test on cortical activity for subject TP05 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0204	**0.0310**
Stimtrack → FCZ2	0.0301	**0.0311**
Audio → FCZ1	0.0216	**0.0296**
Audio → FCZ2	0.0233	**0.0311**
Stimtrack^r → FCZ1	0.0186	0.0102
Stimtrack^r → FCZ2	0.0267	0.0187
Audio^r → FCZ1	0.0182	0.0127
Audio^r → FCZ2	0.0192	0.0181

Table 7.6 The bootstrapping test on cortical activity for subject TP06 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0299	**0.0426**
Stimtrack → FCZ2	0.0276	**0.0557**
Audio → FCZ1	0.0242	**0.0413**
Audio → FCZ2	0.0266	**0.0582**
Stimtrack^r → FCZ1	0.0340	0.0027
Stimtrack^r → FCZ2	0.0288	0.0016
Audio^r → FCZ1	0.0221	-0.0026
Audio^r → FCZ2	0.0183	0.0037

Table 7.7 The bootstrapping test on cortical activity for subject TP07 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0468	-8.1977e-04
Stimtrack → FCZ2	0.0354	-0.0149
Audio → FCZ1	0.0389	0.0056
Audio → FCZ2	0.0363	-0.0043
Stimtrack^r → FCZ1	0.0321	0.0028
Stimtrack^r → FCZ2	0.0278	0.0110
Audio^r → FCZ1	0.0346	0.0067
Audio^r → FCZ2	0.0418	0.0100

Table 7.8 The bootstrapping test on cortical activity for subject TP08 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0263	**0.0383**
Stimtrack → FCZ2	0.0260	0.0048
Audio → FCZ1	0.0218	**0.0433**
Audio → FCZ2	0.0259	0.0141
Stimtrack^r → FCZ1	0.0206	-0.0081
Stimtrack^r → FCZ2	0.0240	6.2468e-05
Audio^r → FCZ1	0.0190	-0.0089
Audio^r → FCZ2	0.0210	-0.0028

Table 7.9 The bootstrapping test on cortical activity for subject TP09 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0307	-0.0297
Stimtrack → FCZ2	0.0369	-0.0343
Audio → FCZ1	0.0400	-0.0325
Audio → FCZ2	0.0393	-0.0422
Stimtrack^r → FCZ1	0.0359	-0.0030
Stimtrack^r → FCZ2	0.0340	0.0021
Audio^r → FCZ1	0.0283	-0.0072
Audio^r → FCZ2	0.0247	-0.0017

Table 7.10 The bootstrapping test on cortical activity for subject TP10 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0151	**0.0211**
Stimtrack → FCZ2	0.0221	0.0215
Audio → FCZ1	0.0114	**0.0199**
Audio → FCZ2	0.0158	**0.0214**
Stimtrack^r → FCZ1	0.0199	0.0060
Stimtrack^r → FCZ2	0.0175	0.0078
Audio^r → FCZ1	0.0130	0.0045
Audio^r → FCZ2	0.0215	-0.0049

Table 7.11 The bootstrapping test on cortical activity for subject TP11 when the outliers have been ignored and the input signals were rectified.

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0164	**0.0299**
Stimtrack → FCZ2	0.0158	**0.0243**
Audio → FCZ1	0.0163	**0.0263**
Audio → FCZ2	0.0173	**0.0294**
Stimtrack^r → FCZ1	0.0152	0.0050
Stimtrack^r → FCZ2	0.0175	2.8603e-04
Audio^r → FCZ1	0.0175	0.0065
Audio^r → FCZ2	0.0163	-0.0055

Table 7.12 The bootstrapping test on subcortical activity for subject TP01 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0090	**0.0234**
Stimtrack → FCZ2	0.0062	**0.0336**
Audio → FCZ1	0.0082	**0.0344**
Audio → FCZ2	0.0068	**0.0338**
Stimtrack^r → FCZ1	0.0080	-0.0061
Stimtrack^r → FCZ2	0.0065	-0.0037
Audio^r → FCZ1	0.0103	-0.0048
Audio^r → FCZ2	0.0083	0.0048

Table 7.13 The bootstrapping test on subcortical activity for subject TP02 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0114	-0.0094
Stimtrack → FCZ2	0.0119	-0.0219
Audio → FCZ1	0.0085	-0.0097
Audio → FCZ2	0.0085	-0.0261
Stimtrack^r → FCZ1	0.0115	-0.0026
Stimtrack^r → FCZ2	0.0090	-0.0025
Audio^r → FCZ1	0.0113	-0.0032
Audio^r → FCZ2	0.0126	-0.0028

B Subcortical responses

Table 7.14 The bootstrapping test on subcortical activity for subject TP03 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0054	0.0041
Stimtrack → FCZ2	0.0058	0.0051
Audio → FCZ1	0.0057	-0.0018
Audio → FCZ2	0.0065	-3.9607e-4
Stimtrack^r → FCZ1	0.0055	-0.0073
Stimtrack^r → FCZ2	0.0070	-0.0018
Audio^r → FCZ1	0.0067	-0.0074
Audio^r → FCZ2	0.0062	-0.0036

Table 7.15 The bootstrapping test on subcortical activity for subject TP04 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0192	-0.0143
Stimtrack → FCZ2	0.0146	9.6005e-04
Audio → FCZ1	0.0116	-0.0117
Audio → FCZ2	0.0117	4.0212e-04
Stimtrack^r → FCZ1	0.0133	0.0050
Stimtrack^r → FCZ2	0.0146	0.0065
Audio^r → FCZ1	0.0122	0.0021
Audio^r → FCZ2	0.0102	-0.0051

Table 7.16 The bootstrapping test on subcortical activity for subject TP05 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0107	-0.0103
Stimtrack → FCZ2	0.0160	-0.0103
Audio → FCZ1	0.0123	0.0063
Audio → FCZ2	0.0128	-0.0067
Stimtrack^r → FCZ1	0.0123	-0.0156
Stimtrack^r → FCZ2	0.0166	-0.0146
Audio^r → FCZ1	0.0110	0.0064
Audio^r → FCZ2	0.0125	-0.0122

Table 7.17 The bootstrapping test on subcortical activity for subject TP06 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0128	-0.0030
Stimtrack → FCZ2	0.0114	-0.0101
Audio → FCZ1	0.0106	-0.0011
Audio → FCZ2	0.0123	-0.0096
Stimtrack^r → FCZ1	0.0096	-0.0041
Stimtrack^r → FCZ2	0.0083	-0.0055
Audio^r → FCZ1	0.0102	6.2404e-04
Audio^r → FCZ2	0.0077	-0.0085

Table 7.18 The bootstrapping test on subcortical activity for subject TP07 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0168	-0.0216
Stimtrack → FCZ2	0.0162	-0.0124
Audio → FCZ1	0.0165	-0.0347
Audio → FCZ2	0.0149	-0.0217
Stimtrack^r → FCZ1	0.0143	-0.0055
Stimtrack^r → FCZ2	0.0174	-0.0130
Audio^r → FCZ1	0.0168	-0.0160
Audio^r → FCZ2	0.0164	-0.0132

Table 7.19 The bootstrapping test on subcortical activity for subject TP08 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0075	-0.0078
Stimtrack → FCZ2	0.0089	-0.0142
Audio → FCZ1	0.0076	**0.0101**
Audio → FCZ2	0.0096	0.0096
Stimtrack^r → FCZ1	0.0057	0.0019
Stimtrack^r → FCZ2	0.0093	0.0023
Audio^r → FCZ1	0.0068	0.0029
Audio^r → FCZ2	0.0089	8.9895e-5

Table 7.20 The bootstrapping test on subcortical activity for subject TP09 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0260	-0.0068
Stimtrack → FCZ2	0.0180	-8.5768e-4
Audio → FCZ1	0.0211	-0.0070
Audio → FCZ2	0.0224	-0.0021
Stimtrack^r → FCZ1	0.0309	-5.7596e-4
Stimtrack^r → FCZ2	0.0234	0.0015
Audio^r → FCZ1	0.0181	-2.4524e-4
Audio^r → FCZ2	0.0284	0.0062

Table 7.21 The bootstrapping test on subcortical activity for subject TP10 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0072	-0.0144
Stimtrack → FCZ2	0.0094	**0.0202**
Audio → FCZ1	0.0070	-0.0165
Audio → FCZ2	0.0098	-0.0235
Stimtrack^r → FCZ1	0.0074	-5.4464e-04
Stimtrack^r → FCZ2	0.0101	-0.0056
Audio^r → FCZ1	0.0073	-0.0014
Audio^r → FCZ2	0.0082	-0.0032

Table 7.22 The bootstrapping test on subcortical activity for subject TP11 when the outliers have been ignored

System	Threshold	estimation accuracy
Stimtrack → FCZ1	0.0092	0.0023
Stimtrack → FCZ2	0.0091	0.0019
Audio → FCZ1	0.0093	-1.8101e-4
Audio → FCZ2	0.0124	0.0041
Stimtrack^r → FCZ1	0.0087	-0.0057
Stimtrack^r → FCZ2	0.0104	-0.0035
Audio^r → FCZ1	0.0081	-0.0053
Audio^r → FCZ2	0.0095	0.0032

Bibliography

Aljarboa, G. S., S. L. Bell, and D. M. Simpson (2022). "Detecting cortical responses to continuous running speech using eeg data from only one channel". *International Journal of Audiology* **0**:0. PMID: 35152811, pp. 1–10. DOI: 10.1080/14992027.2022.2035832. eprint: https://doi.org/10.1080/14992027.2022.2035832. URL: https://doi.org/10.1080/14992027.2022.2035832.

Blom, G., J. Enger, G. Englund, J. Grandell, and L. Holst (2017). *Sannolikhetsteori och statistikteori med tillämpningar*. Sjunde upplagan. Studentlitteratur AB, Lund. ISBN: 9789144123561.

David M. Green, J. A. S. (1966). *Signal detection theory and psychophysics*.

de Cheveigné, A. and I. Nelken (2019). "Filters: when, why, and how (not) to use them". *Neuron* **102**:2, pp. 280–293. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2019.02.039. URL: https://www.sciencedirect.com/science/article/pii/S0896627319301746.

Devasahayam, S. R. (2000). *Signals and systems in biomedical engineering : signal processing and physiological systems modeling*. Kluwer Academic/Plenum Publishers, New York. ISBN: 0-306-46391-1 ;

Fridman, L. (2015). *Fast cross correlation and time series synchronization in python*. URL: https://lexfridman.com/fast-cross-correlation-and-time-series-synchronization-in-python/.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent component analysis*. Wiley, New York. ISBN: 047140540X.

Laguna, P. and L. Sörnmo (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications [Elektronisk resurs]*. Academic Press.

Lalor, E. C. and J. J. Foxe (2010). "Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution". *European Journal of Neuroscience* **31**:1, pp. 189–193. DOI: https://doi.org/10.1111/j.1460-9568.2009.07055.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2009.07055.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-9568.2009.07055.x.

Lalor, E. C., B. A. Pearlmutter, R. B. Reilly, G. McDarby, and J. J. Foxe (2006). "The vespa: a method for the rapid estimation of a visual evoked potential". *NeuroImage* **32**:4, pp. 1549–1561. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2006.05.054. URL: https://www.sciencedirect.com/science/article/pii/S1053811906006434.

Lalor, E. C., A. J. Power, R. B. Reilly, and J. J. Foxe (2009). "Resolving precise temporal processing properties of the auditory system using continuous stimuli". *Journal of Neurophysiology* **102**:1. PMID: 19439675, pp. 349–359. DOI: 10.1152/jn.90896.2008. eprint: https://doi.org/10.1152/jn.90896.2008. URL: https://doi.org/10.1152/jn.90896.2008.

Lalor, E. C., S. Yeap, R. B. Reilly, B. A. Pearlmutter, and J. J. Foxe (2008). "Dissecting the cellular contributions to early visual sensory processing deficits in schizophrenia using the vespa evoked response". *Schizophrenia Research* **98**:1, pp. 256–264. ISSN: 0920-9964. DOI: https://doi.org/10.1016/j.schres.2007.09.037. URL: https://www.sciencedirect.com/science/article/pii/S0920996407004021.

Lindgren, G., H. Rootzén, and M. Sandsten (2014). *Stationary stochastic processes for scientists and engineers*. CRC Press, Boca Raton. ISBN: 9781466586185.

Luck, S. J. and E. S. Kappenman (2012). *The Oxford handbook of event-related potential components*. Oxford University Press, Oxford. ISBN: 9780195374148.

Lv, J., D. M. Simpson, and S. L. Bell (2007). "Objective detection of evoked potentials using a bootstrap technique". *Medical Engineering Physics* **29**:2, pp. 191–198. ISSN: 1350-4533. DOI: https://doi.org/10.1016/j.medengphy.2006.03.001. URL: https://www.sciencedirect.com/science/article/pii/S1350453306000506.

Maddox, R. K. and A. K. C. Lee (2018). "Auditory brainstem responses to continuous natural speech in human listeners". *eNeuro* **5**:1. DOI: 10.1523/ENEURO.0441-17.2018. eprint: https://www.eneuro.org/content/5/1/ENEURO.0441-17.2018.full.pdf. URL: https://www.eneuro.org/content/5/1/ENEURO.0441-17.2018.

MATLAB (n.d.). *Corrcoef*. URL: https://se.mathworks.com/help/matlab/ref/corrcoef.html.

Mitra, P. and H. Bokil (2008). *Observed brain dynamics*. Oxford University Press, New York. ISBN: 9780195178081.

Nousak, J. and D. Stapells (2009). “Auditory brainstem and middle latency responses to 1 khz tones in noise-masked normally-hearing and sensorineurally hearing-impaired adults”. *International Journal of Audiology* **44**, pp. 331–344. DOI: 10.1080/14992020500060891.

Oxley, B. C. (2020). *File:international 10-20 system for eeg-mcn.png*. URL: https://commons.wikimedia.org/w/index.php?curid=86731792.

WHO (2010). *Newborn and infant hearing screening: current issues and guiding principles for action*. World Health Organization, 38 p.

Wikipedia (n.d.). *Tikhonov regularization*. URL: https://en.wikipedia.org/wiki/Tikhonov_regularization.

Lund University
Department of Automatic Control
Box 118
SE-221 00 Lund Sweden

Document name
MASTER'S THESIS

Date of issue
June 2022

Document Number
TFRT-6177

Author(s)
Julia Adlercreutz

Supervisor
Emina Alickovic, Eriksholm Research Center
Martin Skoglund, Eriksholm Research Center
Hamish Innes-Brown, Eriksholm Research Center
Bo Bernhardsson, Dept. of Automatic Control, Lund University, Sweden
Maria Sandsten, Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden (examiner)

Title and subtitle
Brainstem response estimation using continuous sound
A feasibility study

Abstract

Hearing loss is a complicated phenomena which does not only vary from person to person, but also, can change characteristics during the day. Despite this, hearing aids today are fitted only occasionally and thus only capture the slow changes in the hearing loss. In order for a hearing aid to continuously adapt to a subject's hearing loss it has to be able to gauge the users hearing threshold. One way of measuring the hearing threshold is by examining the auditory brainstem response (ABR).

The problem with measuring the ABR today is that it has to be measured as the response to a short sound that is repeated thousands of times. This masters thesis investigates a new method of estimating the brainstem's response to continuous sound. This new paradigm builds on the assumption that the brainstem response corresponds to an impulse response to a system that takes the heard audio as input, and gives the EEG recording as output.

This thesis explores how well this new paradigm works on hearing impaired people that use hearing aids. It verifies that the method works for finding the impulse responses that resemble the cortical response, which is a stronger and slower response. The method was however not successful when it came to finding the subcortical response. A possible reason for this is that a lot of the data needed to be removed due to outliers.

Keywords

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of pages
1-71

Recipient's notes

Security classification

http://www.control.lth.se/publications/