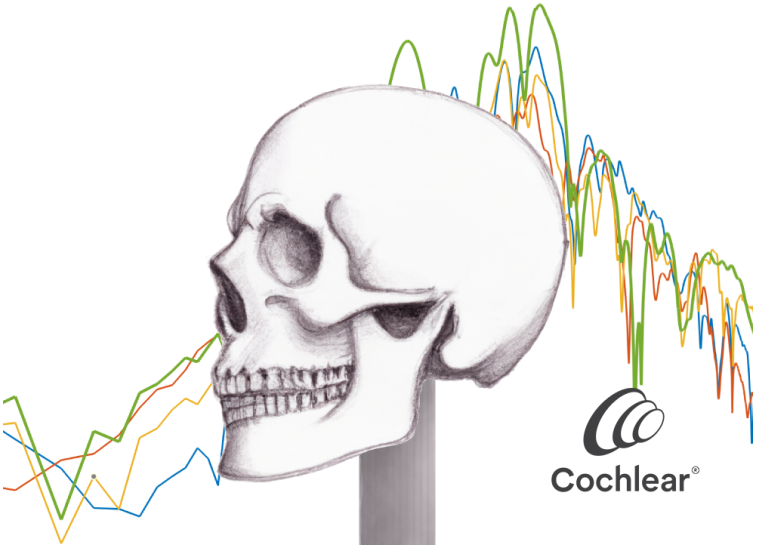
CHALMERS
UNIVERSITY OF TECHNOLOGY



Estimating Bone Conduction Hearing Perception Using Three-Dimensional Vibration in a Head Simulator

Development and Implementation of a Real-Time Combination Model Enabling Realistic Bone Conduction Hearing Estimation

Master's thesis in Biomedical Engineering

Anton Husmark
Erik Håkansson

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

Master's thesis 2023

Estimating Bone Conduction Hearing Perception Using Three-Dimensional Vibration in a Head Simulator

Development and Implementation of a Real-Time Combination Model Enabling Realistic Bone Conduction Hearing Estimation

Anton Husmark
Erik Håkansson

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2023

Estimating Bone Conduction Hearing Perception Using Three-Dimensional Vibration in a Head Simulator
Development and Implementation of a Real-Time Combination Model Enabling Realistic Bone Conduction Hearing Estimation
Anton Husmark and Erik Håkansson

Supervisor: Armin Azhirnian, Cochlear Bone Achored Solutions AB
Examiner: Sabine Reinfeldt, Department of Electrical Engineering, Chalmers

Master's Thesis 2023
Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A sketch of a head simulator with measured vibration components together with the output from the proposed combination model in the background.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Abstract

Assessing the performance of bone conduction (BC) hearing aids is crucial during development. At Cochlear Bone Anchored Solutions AB, one prominent method of accomplishing this is by listening to the hearing aids via head simulators, which are artificial heads equipped with accelerometers as a representation of the cochleae. A new adaptation of a head simulator that uses three-axis accelerometers and has anatomy based on a magnetic resonance imaging scan of a living subject has been developed. There are several indications in the literature that suggest measuring three-dimensional vibration of the cochlea correlates better with BC hearing than one-dimensional vibration, thus suggesting an improvement when having three-axis accelerometers since this enables the usage of all the spatial components. Therefore, this project aims to investigate how to combine three-axis vibration data into a realistic sound representation and how this can be implemented in real-time listening of BC hearing aids on the head simulator.

A proposed combination model is determined through system identification with the vibration data from the head simulator and BC hearing thresholds from the subject at different positions as inputs. The correlated output is deemed to be the corresponding air conduction hearing threshold to correlate three-dimensional vibration to the experienced hearing. The combination model is implemented through finite impulse response (FIR) filters obtained from the system identification, which is analyzed in an offline environment and compared with existing algorithms. Due to the characteristics of the model, it is implementable in digital signal processing (DSP) hardware which means audio signals can be processed in real time where each component from the accelerometer is filtered and then added together.

The FIR filters show a possible way of combining the three spatial components. Simulation of the signal processing of an audio file shows reasonable sound quality. It also indicates a closer connection between vibration and hearing thresholds for signals containing information from all three spatial components, compared to studying one axis at a time. Simulating signals through DSP hardware with the implemented combination filters in real time also shows non-noticeable latency.

These findings indicate that the proposed model accomplishes the aim of combining three-dimensional data into a listenable audio signal. The model is however sensitive to both how the data is acquired, such as stimulation positions used for the hearing thresholds, and the quality of the data in general, such as how good of a representation the head simulator is of the subject's real head. It is also difficult to objectively assess how realistic the sound quality is. However, literature and simulations indicate that the proposed combination model might produce a more realistic representation of BC hearing in a real-time scenario compared to only using one spatial component.

Keywords: Hearing perception, Bone conduction, Bone conduction devices, Three-dimensional vibration, Head simulator, Real-time audio processing, Bone conduction hearing simulation, Digital signal processing

Acknowledgements

We would like to thank Cochlear Bone Anchored Solutions for making this Master Thesis possible. There have been a lot of people involved throughout the project, and we have many to thank. First, we want to thank the audiologist team. Especially thanks to Jenny who helped us with consultation in regards to the audiometric measurements and also helped us to perform the necessary tests that our model revolved around. We also want to thank Jona, who helped us in the Sound Room and consultation early in the project. A big thank you goes out to the TD department where we conducted the thesis work. Henrik helped us a lot in this project in many ways and also helped us feel very welcome around the office space which we are grateful for. We want to especially thank Armin, our supervisor, for the guidance and support throughout the project. We felt as if Armin trusted our process and let us really think outside the box whilst providing guidance and we are really happy with how to workflow of this dynamic worked out. Our examiner Sabine also helped us with the process and we want to thank her for also helping us understand the field better and also for her curiosity.

Last but not least we want to thank our families and friends for their support, not only throughout this project but throughout the years at Chalmers. Thank you!

Anton Husmark and Erik Håkansson, Gothenburg, June 2023

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AC	Air Conduction
ADC	Analog to Digital Converter
BC	Bone Conduction
BCD	Bone Conduction Devices
BNC	Bayonet Neill–Concelman
CPU	Central Processing Unit
DAC	Digital to Analog Converter
DSP	Digital Signal Processing
FIR	Finite Impulse Response
FRF	Frequency Response Function
LDV	Laser Doppler Vibrometer
LTI	Linear Time-Invariant
MISO	Multiple Input Single Output
MRI	Magnetic Resonance Imaging
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
SSD	Single-Sided Deafness
3D	Three-Dimensional
WHO	World Health Organization

Nomenclature

Below is the nomenclature of indices, parameters, and variables that have been used throughout this thesis.

Indices

i	Index for frequency component.
m	Index for cartesian component. Either x, y, or z.
p	Index for the position of stimulation.
s	Index showing which side. Either left, L, or right, R.

Parameters

H_m	Transfer function for the combination model for component m.
$G_{m,p}$	Transfer function for the m-component in *HeadSim 2* at position p.
$T_{BC,p}$	Measured BC threshold at position p.
$T_{AC,p}$	Measured AC threshold at position p.
$A_{m,p}$	Accelerometer reading for m-component corresponding to stimulating with the measured BC threshold for position p.
X_s	X-component accelerometer reading at the side s given an arbitrary sound input.
Y_s	Y-component accelerometer reading at the side s given an arbitrary sound input.
Z_s	Z-component accelerometer reading at the side s given an arbitrary sound input.

Variables

ω	Frequency
t	Time

Contents

List of Figures

1
Introduction

Hearing is an important sense that allows for the most used form of in-person communication. While dependent on the degree of hearing loss, World Health Organization (WHO) states that the impact of hearing loss can affect the life of an individual greatly [1]. This can manifest itself as losing the ability to communicate with others, delayed language development in children, a larger degree of social isolation, and something that is perhaps more common among elderly people; loneliness and frustration.

Furthermore, hearing loss that requires some sort of rehabilitation affects a large percentage of the population, especially in developing parts of the world. WHO estimates that around 5% of the population, 430 million people, suffer from hearing loss that requires rehabilitation [2]. This number is also projected to grow up to 10% of the world population by the year 2050.

Cochlear is an Australian company that specializes in hearing implants [3]. The cochlear implant, which is one of the most known implantable hearing aid, was developed by Cochlear and is used all over the world. In Sweden, a part of Cochlear called Cochlear Bone Anchored Solution has been developing acoustic hearing solutions. The *Baha*® and *Osia*® system are two implantable bone anchored hearing aids developed and produced, which helps people with certain hearing losses hear again.

1.1 Background

Hearing is usually experienced from sound waves that travel through the air and enters the ear canal [4]. The pressure from these waves will vibrate the eardrum, which then transfers the vibration to the malleus, incus, and stapes. This will cause pressure in the fluid of the cochlea, which is filled with hair cells, and when the hair moves back and forth it will trigger a nerve response in the auditory nerve and this is what will be experienced in the brain as sound. This physiological pathway is called air conduction (AC) hearing. The experienced sound also comes from vibrations through the skull bone and other tissues as well and this is called bone conduction (BC) hearing, which is in normal cases the most noticeable when it comes to hearing one's own voice [5]. The physiological pathway for BC hearing ends in the same place as AC hearing, which is in the cochlea. However, the path to the cochlea differs somewhat, and it generally involves vibrations through the skull bones, cartilage, and soft tissues. In the case of hearing one's own voice, the vibration from the vocal cords spread via the soft tissue to the skull bone and thus

produces a hearing response via BC hearing, as well as the regular pathway through AC hearing.

Hearing loss can be caused by various conditions and is also something that can be present at birth, which then often is related to genetic factors [2]. The most common source of hearing loss is related to degradation which is common among the elderly. Other causes, which are more or less common depending on the age of the affected person, include being exposed to loud sounds, infections, chronic diseases, and more. If the hearing loss is severe to the point that barely any sound can be perceived, it is called deafness. Hearing loss can manifest itself in three different ways, namely conductive loss, sensorineural loss, and mixed loss [6]. Conductive loss means that the outer and middle ear is not properly conducting sound to the inner ear, and sensorineural relates to that either the inner ear or nerve pathway between the inner ear and brain is damaged. Mixed loss, as the name suggests, is a mix of conductive and sensorineural loss. Single-sided deafness (SSD) is a severe sensorineural loss in one of the ears. Dependent on the type of hearing loss, rehabilitation differs.

When it comes to conductive hearing loss, AC hearing is compromised [5]. This means that the air pressure from the sound waves in the ear canal does not trigger some of the mechanisms that in the end would trigger a nerve response. In these cases, BC hearing can be used as a replacement for the major contributor to the perceived sound, since the sound waves do not have the same pathway and thus are not necessarily affected by the pathology or defect that is affecting the AC hearing. However, without any aids, the vibrations induced in the skull from airborne sound waves are not strong enough to enable sufficient hearing in a normal setting.

Thus, in order to for BC hearing to be used as a replacement for AC hearing when the latter is faulty, some type of transducer responsible for amplifying vibrations reaching the cochlea needs to be implemented [5]. Such a transducer makes it possible to exploit BC hearing for not only conductive hearing loss but also for mixed hearing loss and SSD. The transducer can either be pressed to the skin, attached directly to the skull, or implanted underneath the skin. The transducer is a key component in bone conduction devices (BCD) [7]. No matter the method of attachment, the BCDs all follow the principle of amplifying airborne sound waves into vibrations in the skull, which will then be transmitted and is in the end what is perceived as BC hearing.

One thing of importance when developing hearing aids, is to be able to review the devices' audibility in an objective manner for an individual patient. When it comes to AC hearing aids, this is usually not a problem [8]. However, for BCDs, this can a more difficult challenge. One standard way of conducting tests is with a skull simulator [9]. The skull simulator is in essence a simple version of the skull in the form of a mass that simulates the mechanical impedance from the skull from the transducer. While effective for some types of measurements, one study mentions that this does not accurately represent a real skull and fails to measure feedback loops [10]. This study also calculated mechanical impedance values of real skulls and these values this not necessarily correlate with the skull simulator. Because of these shortcomings with the skull simulator, Cochlear Bone Anchored Solutions AB has developed a head simulator, *HeadSim 1*, with the impedance values from said study

[11]. *HeadSim 1* consists of a sphere filled with surrogate material, and with the possibility to attach artificial skin. This is an effort to bridge the gap between bench tests and clinical testing, and it can produce results more in line with the clinical data compared to the skull simulator. *HeadSim 1* uses a one-axis accelerometer for measurements of the vibrations.

However, *HeadSim 1* is not able to exactly represent the skull behavior and also degrades over time to some extent. This caused the development of a new head simulator, *HeadSim 2*, where the material and anatomy should mimic that of a real head more closely [11]. *HeadSim 2* is a 3D printed skull structure based on a magnetic resonance imaging (MRI) scan of a living subject. Developing a more advanced head simulator is also interesting from the standpoint that much research today is performed on cadaver heads, which provides good results but is inconvenient for plenty of reasons. In *HeadSim 2*, three-axis accelerometers are placed on both sides of the head at the locations close to the cochleae. Measuring vibration at a location close to the cochlea has been shown to produce better results [12]. Also, BC hearing perception has been shown to be connected to the vibration of the cochlea, although there are less correlation on an individual level [13]. Thus, having the accelerometers placed in a more realistic head simulator is thought to be able to be a better testing environment for BCDs in order to produce a more realistic sound.

Listening to vibration from three axes as opposed to just one is hypothesized to provide a more realistic sound in one article [14]. In this article, the authors investigated how the three components should be added together to produce a sound that is more closely connected to the perceived hearing, but this has never been implemented in a real-time setting.

1.2 Aim

The aim of the project is to enable a better possibility to test hardware for bone-anchored hearing aids by further developing *HeadSim 2*. This is done by developing a way to listen to the vibrations from the head simulator that is representative and realistic with regard to how a person would perceive the sounds from the hearing aids. The aim is to develop a model using two 3-axis accelerometers to measure the vibration at the cochleae locations and produce a stereo sound. The aim is also to have the proposed model implemented, ready to use in physical hardware.

1.3 Objectives

The main objective is to develop a method and a model that translates the two 3-axis accelerometer data to form a stereo sound output. This objective also involves researching how a realistic-sounding BC hearing estimation could be achieved in the head simulator. The main objective is broken down into parts.

- To design a model structure that correlates the hearing perception given a vibration input of a bone-conducting hearing aid.

- To implement a method in a programming and computing platform that identifies and quantifies every aspect of the model.
- To implement a proof of concept an analytical tool that enables simulation of the head simulator and the proposed model. This means inputting prerecorded signals and is thus an offline signal processing step.
- To implement the signal processing chain in hardware which should accomplish the listening in real-time.

1.4 Demarcation

Following the aim of this project, there is no need to research and describe the underlying mechanics correlating cochleae vibrations and perceived hearing, and thus no effort will be made to describe this behavior in general terms. Also, the calculated model for vibration combination is being developed for the specific head simulator *HeadSim 2*. This means that the model and method could be specific to this head simulator and investigations regarding how it performs in a general case are not conducted. The model is however discussed in the context of applicability.

Since the number of head simulators based on living subjects is limited to one, it will not be investigated how any determined method and model will deviate between head simulators. Thus, the proposed method and model will not be analyzed through statistical certainty. However, since the exact mechanism is not investigated and instead realistic sound should be produced for this specific skull, this is assumed to be acceptable in this project.

In terms of hardware, such as with the accelerometers and hardware used in real-time processing, it is not investigated if other components would provide different results, i.e. the available hardware is used. How computationally efficient different methods and algorithms are also not investigated. This means that it is not analyzed if changes in these algorithms or methods would minimize latency or enable the use of other available algorithms.

2
Theory

There are several aspects that are relevant to have an understanding of in order to develop a method that accomplishes the aim of modeling a combination of 3-axis accelerometers. One of these aspects is having a general understanding of the vibration pathway through the skull and how this relates to BC hearing. How these vibrations are introduced into the skull with actuators is also relevant. Knowledge regarding *HeadSim 2*, such as materials and accelerometer, is also central. In order to create a model, general knowledge of the mathematics surrounding audio signals is presented. Also, in order to connect the model to the BC hearing experience, audiometric measurements and the current research in regard to BC sound perception are relevant. Finally, in order to implement a model in real-time, it is important to understand how to process signals with minimal latency and noise with hardware.

2.1 Vibration Pathway Through the Skull in BC Hearing

While the main principle of BC hearing relates to different types of tissue vibrating and thus transmitting the vibrations to the Cochlea and thus providing hearing stimuli, the vibration pathway in the skull is quite complex [15]. This is because the skull consists of several different bones with different shapes and thicknesses, cartilage, and soft tissues such as the brain and skin, which all got different vibration and attenuation characteristics. The different bones are connected with sutures, which also affects vibration transmission. All of these factors affect the resulting cochlea vibration based on the position where the vibration is introduced since a different pathway means a difference in both distances traveled and tissues traversed.

Furthermore, the frequency of the vibration has a great effect on the way it is transmitted through the skull. At lower frequencies, between 150 and 400 Hz, the skull behaves similarly to a rigid body [15]. This is because this frequency range is below the resonance frequency mechanical impedance point of the skull. The rigid body behavior means that the whole skull moves with the vibration. For example, the direction a vibration is introduced will dominate the vibration of the skull, and thus also the cochlea. At higher frequencies, the skull behaves more as a mass-spring system. Depending on the frequency range, the type of system varies. For example, in between the frequency ranges of 150-400 and 1000 Hz, large parts of the skull move in phase. Above 1000 Hz, the mass-spring behavior is dominated by

wave transmission. Because of this, higher frequencies tend not to be transmitted as strongly as lower frequencies over a distance, since transversing through a new medium is affected by wave transmission.

2.2 Actuators in BCD

The actuator is responsible for introducing motion in a system, and in the case of BCD, this can be accomplished in a couple of different ways. The *Baha*® and *Osia*® system use an electromagnetic motor and a piezoelectric element respectively in order to introduce vibration to the system [16].

The resonance frequency of an actuator is relevant in acoustic hearing aids since it affects the force output. The resonance frequency is different deepening on the type of device. Even within one type of actuator, it can vary greatly deepening on design factors. For an electromagnetic transducer, such as in the *Baha*®, the mechanical resonance depends on the inertial mass's motion in relation to the skull, in combination with any suspensions in the system [17]. The resonance frequency/frequencies are often designed to fall below or above frequency/frequencies of interest. An example can be seen in figure 2.1, where the force output from an arbitrary actuator developed for use in a *Baha*® system can be seen. In this particular actuator, resonance frequencies can be seen at around 800 Hz and just below 10 000 Hz.

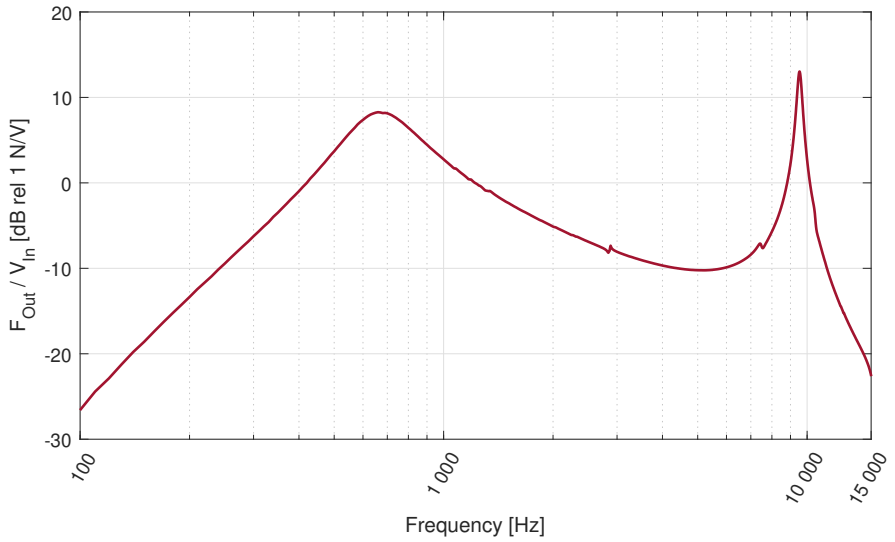


Figure 2.1: Frequency response of an actuator developed for use in a *Baha*®-system. It shows force output given a voltage input.

A piezoelectric transducer, such as the *Osia*® system, moves alongside the skull and introduces vibration through a localized bending moment that elastically deforms the skull [16]. This type of transducer does not have the inertial mass and suspension system present in the electromagnetic transducer. Thus, the mechanic resonance does not behave in the same manner.

2.3 HeadSim 2

The materials for *HeadSim 2* are chosen to match the properties of a real skull. This includes *Perma-Gel* as a surrogate material for the brain, which was investigated in a previous master thesis at Cochlear [11]. *Perma-Gel* is a synthetic gel material that is similar to gelatin and is non-aqueous.

While *HeadSim 2* does not have any surrogate skin material, there exist options to attach some type of synthetic skin. This is relevant to include in some cases where proper attenuation and dampening are needed. Options for synthetic skin were also explored in a previous master thesis, where a material manufactured from cured silicon yielded the best result [18]. In the thesis, the authors recommended several layers of this material to produce the best result.

2.3.1 Accelerometers in *HeadSim 2*

The 3-axis accelerometers are placed at positions close to the true cochlea positions. The accelerometer directions are defined as seen in figure 2.2, with the x-direction being front-to-back, the y-direction left-and-right, and the z-axis up-and-down. The accelerometers are the *Dytran model 3293A* [19]. The accelerometers have a sensitivity of 500 mV/g, a weight of 8.8 g, and a resonance frequency above 15 000 Hz. The max transverse sensitivity, which is how much the axes are separated, is 6% or around 24 dB meaning that the axes can at maximum be 24dB apart.

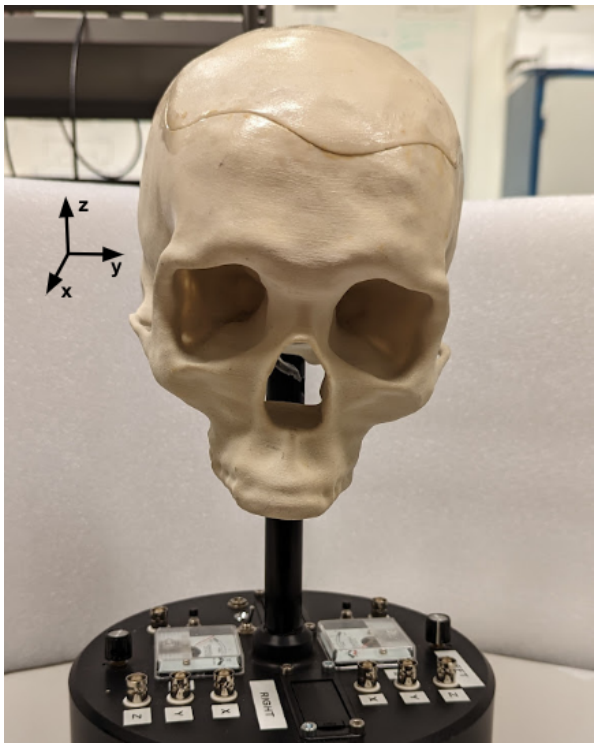


Figure 2.2: A front view of *Headsim 2* showing the orthogonal accelerometer directions.

The skull is mounted on a base where amplifiers for the accelerometers are located, which can then be connected with coaxial cables with Bayonet Neill–Concelman (BNC) connectors to analyze and/or process the signals from the accelerometers. These connections are available for both of the accelerometers located at each side of the skull.

2.3.2 Attachment of Actuators

Actuators can be attached in several ways depending on the type of BC-transducer, as described in section 1.1. For the *Baha®* system, a percutaneous implant is often used. The implant is a *BI300*, which is anchored to the skull. The *Baha® BA400 Abutment* can be attached to the implants where the actuator can be attached. As for the *Osia®* system, a similar system exists, although the implant is transcutaneous [7]. The position of the abutment of these two systems is different, with the *Osia®* system being located closer to the ear canal compared to the *Baha®* system.

HeadSim 2 contains several of the *BI300* implants at *Baha®* position and *Osia®* position on both sides of the head, where it is possible to attach *Baha® BA400 Abutment*. This makes for a convenient testing method of the actuators since this will replicate how the *Baha®* and *Osia®* are attached on an actual patient.

It is also possible to attach actuators with either a *SoftBand*, *HeadBand* or *TestBand*. The *SoftBand* was developed to provide patients that is too young to be considered for the surgery required for the percutaneous implant [20]. As the name suggests, the band that holds the actuator is flexible in order to be comfortable throughout the day. In order to have a good balance between comfort and audio quality, the force of the band is often set to around 2 N. As for the *HeadBand* and *TestBand*, they are similar to the *SoftBand* in the sense that they can be used without abutments and thus do not require any surgery. Compared to the flexible material of *SoftBand*, these options are made from a steel spring band instead. The main difference between the two is that the *TestBand* has a higher force of 5 N, compared to the 2 N of the *HeadBand*. Both of these options can be used for preoperative audiologic evaluation [21]. They produce similar results, however, the higher force of *TestBand* leads to slightly better throughput, at the cost of being less comfortable. In terms of performance, *HeadBand* and *TestBand* perform around 5-20 dB worse in hearing threshold compare to the abutment.

For the attachments that require skin contact, i.e. all of the transcutaneous attachment methods, pressure necrosis can be an issue [22]. Thus, it exits something called a *SoftWear Pad* that relieves the pressure at the contact area with the actuator by distributing the area of contact in a more even manner.

2.4 Modeling

Acoustic or audio signals generally result in linear systems, unless they use psychoacoustic models to compress the signals [23]. Time invariance is also generally approximated and valid for acoustic and audio signal systems. Thus, said systems are modeled as linear and time-invariant (LTI) systems.

The input and output relation of an LTI system can be completely described by its impulse response, $h(t)$ [23]. When the system only has one single input and one single output, the system is called a single-input-single-output (SISO) system. Calculating the output, $y(t)$, from the input, $u(t)$, of such a system can be done through equation 2.1.

$$y(t) = h(t) * u(t) \tag{2.1}$$

The impulse response can be analyzed and determined in the frequency domain. Laplace transform can be used in equation 2.1 resulting in a multiplication instead of convolution between the input and output relation. For most acoustic and audio signal systems, the initial condition is zero, which results in that $s = j\omega$, resulting in the Fourier transform. Thus, using Fourier transform on equation 2.1 results in the equation 2.2.

$$Y(j\omega) = H(j\omega) \cdot U(j\omega) \tag{2.2}$$

$H(j\omega)$ is called the transfer function and relates the input to the output in the frequency domain. Given this relation in the case of a SISO system, it follows that a vector representation can be used to generalize in the case of a multiple-input-single-output (MISO) system [24]. This results in the equation 2.3.

$$Y(j\omega) = \mathbf{H}(j\omega)\mathbf{U}(j\omega) = \begin{bmatrix} H_1(j\omega) & H_2(j\omega) & ... & H_n(j\omega) \end{bmatrix} \begin{bmatrix} U_1(j\omega) \\ U_2(j\omega) \\ ... \\ U_n(j\omega) \end{bmatrix} \tag{2.3}$$

where n is the number of inputs contributing to the output. In system identification of an LTI system, the transfer functions can be seen as unknowns, thus needing at least n measurements to uniquely define the system transfer functions.

2.5 Audiometric Measurements

In audiometry, there are several ways of testing the hearing mechanism in a subject. One of the most prevalent is *pure tone audiometric air conduction testing*, where a pure tone is delivered to the ear via headphones, and then measure the lowest possible intensity in dB which can be detected by the subject 50% of the time [25]. A measurement such as this is called *threshold*. A normal threshold measurement is performed from 250 to 8000 Hz. For diagnostic, it is not practiced to test frequencies above 8000 Hz since it is not necessary to obtain information for the higher frequencies. This is mainly because the frequency range 250-8000 Hz is where the human speech spectrum is located. Performing threshold measurements are possible for both AC and BC hearing. When several of these thresholds are acquired at specific values in the frequency range, it results in an audiogram. AC and BC audiograms are often performed on the same subject in order to assess if a hearing deficiency is due to conductive and/or sensorineural loss. When acquiring BC hearing thresholds, the transducer is placed on the mastoid.

There are several things affecting the uncertainty of an audiogram [26]. This includes physiological factors such as spontaneous activity in the central nervous system, tinnitus, adaptation, and fatigue. Uncertainties also come from psychological factors, which can be in the form of tiredness, motivation, and concentration level. The methodology also affects the uncertainty. Since there are a lot of factors that can affect the results of an audiogram, it is said that even if the same test is retaken in a short time window, it will most likely not be identical to the first test. The standard deviation of the difference between repeated measurements is around 3-7 dB for the AC threshold, and 4-10 dB for the BC hearing threshold [27].

2.5.1 Masking

Masking is something that is implemented on the non-testing ear in audiometric testing when it is possible that testing might produce a hearing response at the non-testing ear [28]. The goal of masking is to introduce an acoustic sound in order to mitigate any sound perception given from the sound propagating to the non-test ear. The masking noise can be chosen to be either broad- or short-band white noise. It can also be located within the speech range of 300-3000 Hz.

The noise needs to be calibrated. This can be done via either dedicated equipment, or the more commonly used method among clinicians which is with biological calibration methods using people [28]. This is done by testing the signal simultaneously as the masking signal, and it is seen as an effective masking if the signal is just barely masked. This can be seen as calculating how much the masking can shift the threshold of the signal in dB. The thresholds are tested at several intensities for the masking noise until these threshold shifts are linear. This method requires several test subjects, often 3-5, where the results of the subjects are averaged. Even though several subjects are needed for this type of calibration, it is seen as a fast way to calibrate the masking noise.

2.5.2 Occlusion Effect

When testing BC hearing for a person with normal hearing, it can be of interest to plug the ear in order to cancel out any air conduction hearing of the actuator that is vibrating. However, this is something that is not always possible, due to something called the occlusion effect [29]. The occlusion effect means that sound appearing in the ear canal via the BC sound is amplified, and thus providing a hearing response via AC hearing in conjunction with the BC hearing. This is something that is the most prevalent at lower frequencies. Due to the occlusion effect, blocking the ear canal during BC simulation affects the threshold, since it can give the impression of better hearing at these frequencies. If there is a desire to block the ear canal, inserting a plug deeper into the ear canal lessens the effect of occlusion [29].

2.5.3 Warble Tones

A warble tone is a tone where the frequency varies within a small interval several times a second. In audiology, it is typically implemented when performing thresholds

for pediatric testing, patients with tinnitus, and when performing the test with speakers instead of headphones when it comes to AC testing [30]. It generally produces similar thresholds as regular pure tone testing. However, some studies suggest that warble tones produce lower hearing thresholds [31]. Despite this, it is generally thought to be possible to substitute pure tones with warble tones in general tests as well, and in some cases, it is preferred by the test subject even when not falling under previously defined cases [32]. In some cases, it has also been reported that warble tones produce fewer false positives compared to pure tone testing, which by extension could mean more effective testing and less exposure for the subject.

At higher frequency testing, one study has shown a greater dispersion between pure tone and warble tone testing [31]. In the study, they found that testing frequencies close to the subject's upper limit with warble tones greatly affected the hearing threshold. The authors attributed this to the fact that a warble tone consists of a frequency band with +/- 5% of the center frequency, and thus the lower part of the tested frequency can be located within the hearing threshold of the subject. In practice, this means that the subject might be wrongly assumed to hear much higher frequencies than in pure tone testing.

2.6 BC Sound Perception

It is, in general, difficult to model the BC sound perception since the vibration behavior of the skull and all the pathways are frequency-dependent making it complex, as explained in an article [15]. The article also explains that it is interesting to study the vibration response of the cochlea and that it is generally thought that a stronger cochlea vibration is connected to a louder BC sound. However, at the time of this article, the exact relation between the two was not clear. The author also mentions that previous studies have shown that the cochlea does vibrate in all spatial dimensions and that no component has been shown to be dominant.

In one study, differences in hearing BC threshold between stimulation positions at the mastoid and close to the mastoid position were investigated [33]. In this study, one-axis vibration measurements on the cochlea of cadaver heads were compared with several subjects hearing thresholds of the same stimulation thresholds. The authors concluded that small changes in stimulation position did not have a significant effect on the hearing threshold.

2.6.1 Cochlea Vibration and Hearing Perception

In one study, the correlation between hearing perception and cochlea vibration was investigated [13]. In this study, a Laser Doppler Vibrometer (LDV) was used to measure the vibration of the cochlea promontory on live subjects in a combination of a BC hearing threshold. The conclusion from this study was that there was a correlation between the vibration curve and the hearing threshold curves if looking at a group level, but not at an individual level. However, the authors mention that only the vibration in the y-direction, with directions according to figure 2.2, was measured. Even though this is where the vibrations are the most dominant for

the chosen position of stimulation, and thus assumed to be the largest contributor to hearing perception, the other directions could provide more information. This might explain the deviation between the individual correlation between vibration and hearing perception.

2.6.2 Using Three-Dimensional Vibration

In terms of how to combine a 3D vector into one vector, there has been a study investigating how to combine the different vibration directions and represent them in one vector with magnitude and phase [14]. This study, which is referred to as "Dobrev et al." here on now in this report, concluded that combining the 3D vector into one vector would be closer to BC hearing than taking any individual spatial contribution. This was achieved with a maximum velocity vector, and it was verified by using the average of multiple hearing thresholds. To make this conclusion, the study used cadaver heads, a *Baha*® transducer, and a 3D LDV. They then analyzed the transcranial attenuation of the maximum velocity vector and correlated this to the threshold differences between the stimulation positions. The algorithm for calculating the maximum velocity vector includes using square roots, logarithmic functions, and trigonometric functions. However, there was no effort to connect the measured cochlea vibration and the hearing threshold for the individuals, since this study was performed on cadaver heads. It was also noted that many of the individual variations that exist when studying one direction at a time were not present to the same extent when looking at the combined vector.

2.7 Digital Signal Processing

Digital Signal Processing (DSP) is a method for realizing signal processing in real-life scenarios. This is practical in the sense that signals can be filtered in real-time with minimal delay, which can be used in many fields of engineering. For example, in the previous installments of a head simulator at Cochlear Bone Anchored Solutions, *HeadSim 1*, DSP hardware has been used in order to process and filter the one-axis vibration signal in various ways in order to listen to BC transducers in real-time. This has been found to work well for the purpose of listening to a head simulator.

2.7.1 Signal Chain in DSP

DSP is based on sampling real-time signals, such as an audio signal, and converting them to a digital signal which can then be processed and manipulated with the digital signal processor [34]. In the DSP chain, the sampling is performed by the analog-to-digital converter (ADC), which can be seen as the first block in figure 2.3.

The ADC also quantizes the signal after the sampling and also encodes the signal to the digital signal after quantization [34]. Sampling the signal means taking samples in a defined time interval of the signal. The reason it is needed is that the analog signal is continuous, and is thus impossible to process since this would require an

infinite amount of memory and processing power. There are some things to take into consideration in terms of the sample rate, i.e. how often a sample is acquired from the signal, such as the principle of Nyquist sampling. This means that the sampling rate needs to be at least twice that of the signal bandwidth. Another way to view this is that a signal needs to be sampled at least twice per period in order to be reconstructed. Having a higher sample rate relates to more processing, and is thus more demanding for a DSP system.

In the middle block seen in figure 2.3, the processing of the signal takes place. When it comes to digital processing of the signal; digital filters are often implemented [35]. These filters can be a combination of delay, summation, and multiplication. A standard central processing unit (CPU) present in ordinary computers can handle the first two of these operations efficiently, but not multiplication, and thus latency is often an issue when performing DSP with such a CPU. This means real-time signal processing becomes an issue. There exists DSP hardware that is specifically built for multiplication, and implementing digital filters in these means better real-time processing performance compared to a regular CPU.

When the signal has been processed, it can then be converted back into a time signal and thus, in the case of an acoustic signal, make it listenable [34]. This is performed in the digital-to-analog converter (DAC) which is seen in figure 2.3.

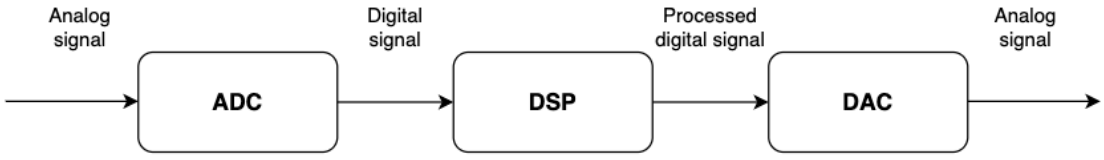


Figure 2.3: DSP block diagram showing the signal processing pathway. An analog signal is converted to digital in the ADC block, then processed in the DSP block, and finally converted back to an analog signal in the DAC block.

2.7.2 DSP in Audio

Audio signals are defined within the audible frequency span, which consists of frequencies between 20 and 20 000 Hz. A constant sampling rate of at least 40 000 Hz is thus needed in order to fulfill Nyquist sampling [34]. While 40 000 Hz is enough to cover the frequency range in theory, the standard when dealing with audio signals is to use a sampling rate of either 44 100 or 48 000 Hz.

When dealing with audio in real-time, latency is seen as an important factor and if is noticeable it will be perceived as a delay between the source of the sound and the output. An acceptable latency level can vary depending on the application, but a latency below 10 ms is usually seen as imperceptible for people [36]. For a DSP system to be able to accomplish the processing with a latency lower than 10 ms, it depends on the power of the system in terms of processing speed, as well as the type of algorithm implemented in the processor.

Noise is something that is generally present in many areas of signal processing, and nonetheless in audio signals. In a DSP chain, this means that not only the signal of interest is present, but also different types of disturbances that interfere the signal to various degrees. Signal-to-noise ratio (SNR) is a concept that explains of much noise there is in relation to the signal, and a system generally aims to have a high SNR for the best performance [23]. The noise floor is another concept that relates to the sum of all the noise existing in the system.

3
Methods

In order to find and implement a model to make the vibration components from *HeadSim 2* listenable in a pair of headphones, multiple aspects are needed to be taken into consideration. The model was required to be implementable in hardware with low latency. Firstly, the structure of the system was determined where it relied on findings in literature and identification of aspects unique to the domain of *HeadSim 2*. From the overall structure, a plan for acquiring the data needed was formed. Lastly, the proposed combination model was implemented in DSP hardware.

3.1 System Structure

The overall structure of the implemented solution relied on both findings in the literature that could bridge the gap between skull vibration and hearing perception, as well as system identification of components unique to this system. The signal pathway of the system included six analog signals from the two three-axis accelerometers as inputs to the model and two channels of analog audio signals as the output. Since the hearing perception for each of the output channels could correlate independently to one cochlea, the signal pathway could be split into two parallel pathways with two models. The block diagram showing the steps involved in listening to the head simulator can be seen in figure 3.1.

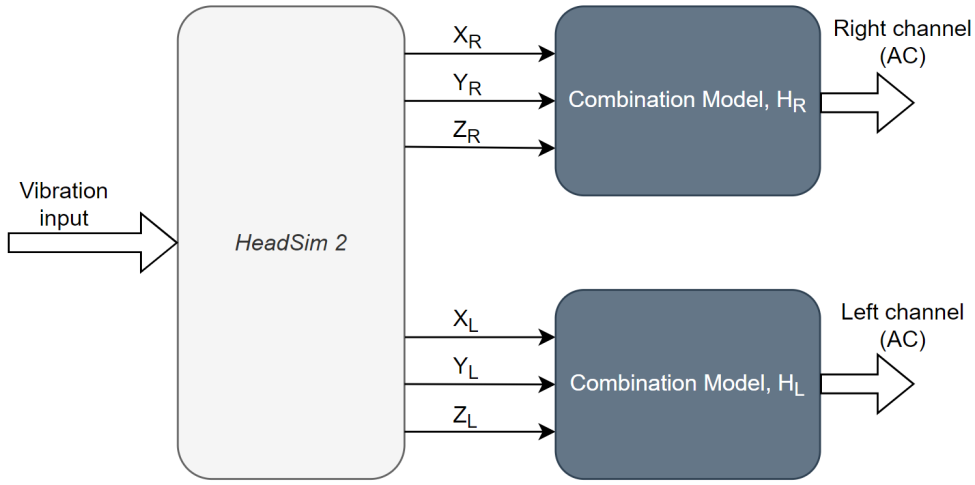


Figure 3.1: A block diagram showing the system enabling a real-time combination of accelerometer readings to represent BC hearing. X, Y, and Z depict the orthogonal accelerometer components for the respective right or left side given an arbitrary sound input.

The block diagram shows that any sound input that is converted into vibrations, e.g. by actuators mounted on the skull, that are propagating through the skull are then measured by the accelerometers. The accelerometer readings are then combined into one single output signal by a combination model. A possible way of combining the spatial components was presented by Dobrev et al. where the authors suggested a method of using the three-dimensional data in the computation of a maximum velocity vector. This was however investigated in cadaver heads and did not make any effort to make the method processable in real-time, suggesting limitations to its use for the aim of this project. Due to the characteristics such as size, weight, and deviations in the placement of the accelerometers, the determination of this combination was done through system identification where the proposed combination model was regarded as a black box. This could also enable capturing of dynamics that are unique to the head simulator and those that might be introduced through the choice of components in the head simulator, and still make it able to match a realistic hearing experience.

One aspect of the combination model is the correlation between AC and BC hearing thresholds. The idea is that AC and BC sound perception is estimated to be similar enough, and thus it should be possible to correlate the two as input and output since the goal is to listen to the BC sound perception as AC sound. Even though this is a simplification, it is deemed to be reasonable for the objective.

The determination of the combination model was done through a hearing threshold matching against the measured vibrations at the cochlea position. Previous investigations, described in section 2.6, have used matching of hearing threshold and cochlear vibration in order to assess the BC hearing perception of the subjects. This can suggest that the input and output from an audiometric test could be used to match the output from the accelerometers to the output of the physical system, i.e.

hearing perception in a live human being. Even though some literature suggests a poor correlation on an individual level when using a one-axis vibration measurement, it can be hypothesized that three-axis measurements can improve this correlation. Because of this, the method of determining the combination model used threshold matching to vibrational data. Since the head simulator is based on an MRI of a specific subject, the specific audiogram of the same subject was used in the system identification method. This determination of the combination model can be seen in figure 3.2.

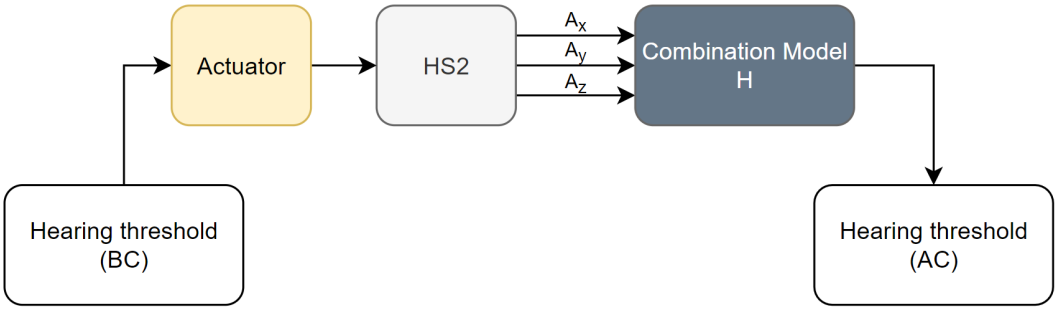


Figure 3.2: Using a known and measurable test case, the model H can be determined. BC thresholds are used as the measured inputs, and the AC thresholds are used as the measured output then the output in the headphones should be the AC thresholds. A_x, A_y, and A_z depict the measured vibration in the direction of the orthogonal components respectively.

3.2 System Identification

The first step of identifying and determining a model converting accelerometer data to headphone input is to find the constraints of the model and what properties the system will have. Since the audio signal system converts sampled accelerometer data and outputs headphone input voltage it can be assumed to be linear. Time invariance is also assumed for such a system since the mapping is not substantially dependent on time. Even though psychoacoustic elements might influence the gathering of data leading up to the determination of the system, the final model is not designed to model any psychoacoustic principles. The model instead strives to make the conversion and summation of the accelerometer data to headphone input. Therefore, the combination model was seen as an LTI system.

Each of the cochlea perceptions can be regarded as two different MISO systems with three sampled accelerometer signals as input and outputting the right or left channel of the headphones respectively. These systems can be described with a vectorized transfer function in accordance with equation 2.3. In a system identification setting, these transfer functions can be seen as unknowns in a system of linear equations. Directly following the MISO transfer function representation, one equation system can be derived for each of the two accelerometers. In order for the solution for the equation system to be unique, it has to have three linearly independent equations

while having three unknowns. This equation system can be seen in equation 3.1,

$$\begin{cases} H_x(j\omega_i)A_{x1}(j\omega_i) + H_y(j\omega_i)A_{y1}(j\omega_i) + H_z(j\omega_i)A_{z1}(j\omega_i) = T_{AC}(j\omega_i) \\ H_x(j\omega_i)A_{x2}(j\omega_i) + H_y(j\omega_i)A_{y2}(j\omega_i) + H_z(j\omega_i)A_{z2}(j\omega_i) = T_{AC}(j\omega_i) \\ H_x(j\omega_i)A_{x3}(j\omega_i) + H_y(j\omega_i)A_{y3}(j\omega_i) + H_z(j\omega_i)A_{z3}(j\omega_i) = T_{AC}(j\omega_i) \end{cases} \tag{3.1}$$

where H_x, H_y and H_z are the sought-after frequency responses. $A_{m,p}$ is the sampled accerelometer input in the frequency domain in m-direction for the different examined positions, p. The index i is denoting the measured frequency response for that particular frequency. Three position measurements are required when solving the equation system and are denoted as 1, 2, or 3. The solution to this linear equation system determined the frequency responses for the different filters uniquely.

In order to acquire all of the information needed to solve the linear system of equations, multiple measurements were needed. This included gathering the frequency response of *HeadSim 2*, $G_{m,p}(j\omega)$, for each axis at the different locations, BC hearing thresholds, $T_{BC,p}(j\omega)$, at the same locations, and the corresponding AC threshold for the measured cochlea, $T_{AC,p}(j\omega)$. The procedure for gathering this data is described in 3.3.

With this data, the linear equation system was solved in MATLAB. The acquired frequency responses of *HeadSim 2* at the different measurement positions were used in conjunction with a sinc interpolation. This was accomplished by truncating the measured frequency responses in the time domain. This could then be used together with the measured input voltage to the actuator was used to find the corresponding accelerometer readings, $A_{m,p}(j\omega)$. This accelerometer reading corresponds to a Cartesian component, m, in *HeadSim 2* when the actuator is stimulating with the same stimuli that resulted in a BC hearing threshold for that position, p. Following this, every accelerometer reading was determined through equation 3.2.

$$A_{m,p}(j\omega) = G_{m,p}(j\omega) \cdot T_{BC,p}(j\omega) \tag{3.2}$$

This is then used as $A_{m,p}$, as seen in equation 3.1, and assembled in a square matrix $\mathbf{A}$. The measured inputs to the used headphones during the air conduction hearing test are assembled into a row vector $\mathbf{T}$. Since the matching output, $\mathbf{T}$, will carry information about both the output in regards to both magnitude and phase, a phase component is added to the measured AC threshold. Thus, $\mathbf{T}$ is carrying information regarding both volume and timing for the frequencies. This phase component is set to the mean of the phase angle for orthogonal accelerometer readings for each of the measured frequencies at each of the positions. This is assumed to be a reasonable suggestion when the phase of the individual components difference is relatively small compared to the differences when changing position. The linear equation solved in MATLAB can be seen in equation 3.3

$$\mathbf{AH} = \mathbf{T} \implies \begin{bmatrix} A_{x1}(j\omega) & A_{y1}(j\omega) & A_{z1}(j\omega) \\ A_{x2}(j\omega) & A_{y2}(j\omega) & A_{z2}(j\omega) \\ A_{x3}(j\omega) & A_{y3}(j\omega) & A_{z3}(j\omega) \end{bmatrix} \begin{bmatrix} H_x(j\omega) \\ H_y(j\omega) \\ H_z(j\omega) \end{bmatrix} = \begin{bmatrix} T_{AC1}(j\omega) \\ T_{AC2}(j\omega) \\ T_{AC3}(j\omega) \end{bmatrix} \tag{3.3}$$

The solutions to the equation systems were further modified in order to result in realizable filters in hardware. The first modification from the determined frequency responses was done to mitigate the sharp fluctuating slopes. This was done by using a moving third-octave band mean. The second modification to the determined filter was to modify the filters to span the full frequency spectrum. This included assigning values to the frequency response between determined values and outside of the measured frequency span. The frequency response value from the lowest measured value is kept down until 25 Hz. Below 25 Hz and above the highest measured frequency an attenuating slope was added. A Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation was used to interpolate between the determined and designed values of the frequency response.

The frequency responses were then transformed to the corresponding impulse response in the time domain by using the Inverse Fast Fourier Transform (IFFT). The impulse responses were shifted, truncated at 1500 taps, and windowed with a Hanning window. This was performed to remove any wrap-around, to make the filters reasonable to realize in real-time signal processing, and to reduce any coefficients that might have a considerable magnitude at the extremes respectively.

By convolving the impulse responses from the FIR filters with the corresponding orthogonal components from the accelerometer readings and summing them together, a single audio signal was acquired. These FIR filters were thus available for both offline and real-time evaluation.

3.3 Measurements

In order to solve the equation system, several measurements were conducted over several of frequencies with the help of an actuator. This included threshold measurements on a test subject, i.e. audiometric tests, and frequency response measurements on *HeadSim 2* at several locations.

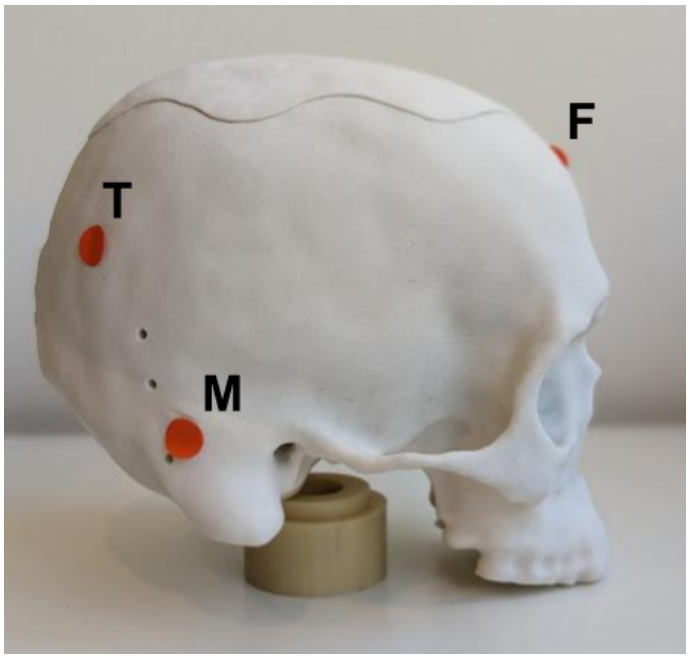
3.3.1 Measurement Setup

Since the combination model estimation relies on measured data, both in accelerometer readings and audiometric tests, it was important that this data consisted of the highest possible quality. This means that the chain producing and propagating the vibration to the cochlear site, both in the head simulator and the test subject, should be as identical as possible. The two conducted measurements, therefore, included the same actuator, actuator positions, attachment method, and skin interface. The measurement of the hearing threshold was also connected to the corresponding measured accelerometer in the head simulator, meaning that only the hearing of the intended cochlea was measured, regardless of the transducer placement during measurement. Only one cochlea was tested in this fashion, and the model for the other cochlea was assumed to have the same thresholds. This was to half the amount of time performing the measurements in order to lower the inconvenience for the test subject. Also, an AC threshold was conducted on both ears where it was assessed that the subject had similar enough hearing thresholds in order to make this assumption.

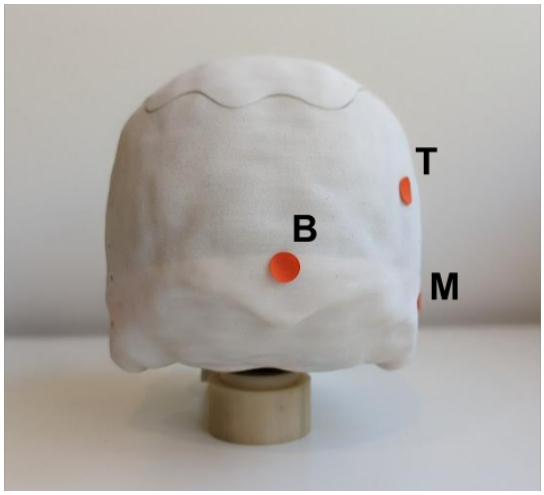
For the measurements, a modified *Baha®* actuator was used in order to properly stimulate frequencies up to 14 000 Hz. Normally, these actuators are not designed to operate in this frequency range, thus it was tested in order to judge that no distortions appeared at the desired frequencies. In order to simulate different positions on both *HeadSim 2* and the test subject, either a soft-band or steel-spring band was a possible solution in order to attach an actuator to the test subject. A steel-spring band in the form of *TestBand* was chosen because of its beneficial properties in terms of throughput. However, since several thresholds were gathered, combined with the fact that the *TestBand* is notorious for being experienced as uncomfortable because of the higher force of the band, *SoftWear Pads* were implemented in order to relieve some of the pressure.

3.3.2 Stimulation Positions

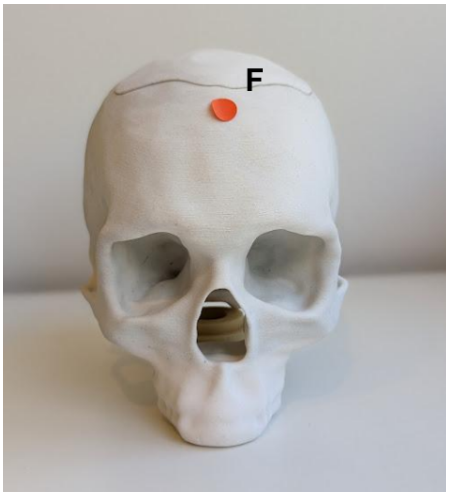
In terms of the used measurements, four bone conduction thresholds at different positions and one air conduction test were performed. This is because three bone-conduction tests and one air-conduction test is the minimum requirement to solve the proposed combination model described in section 3.2, and one more measurement was performed to verify the created model. Since the literature suggest that the vibration response and thus hearing threshold can be similar for stimulation positions, the positions for this setup were chosen in such a way that the complete model of the vibration to perceived sound was as complete as possible. Thus, the positions chosen were the mastoid, the back of the head, the forehead, and a position in between the neck and the mastoid, which can be seen in figure 3.3. From here on now in the report, these positions will be called mastoid, back, frontal, and triangle positions, which in the figure 3.3 are marked as M, B, F, and T respectively.



(a) Sideview of the skull with triangle, frontal, and mastoid positions marked out.



(b) Backview of the skull with back, triangle, and mastoid positions marked out.



(c) Frontview of the skull with the frontal position marked out.

Figure 3.3: The 3D printed skull with triangle position T, frontal position F, back position B, and mastoid position M marked out.

The literature suggests that having BC threshold obtained at positions close to each other produces similar results. In conjunction with the general measurement uncertainties present in audiometry, this indicates that it is reasonable to have the stimulation position far apart since it otherwise would have been difficult to asses whether differences in obtained data come from the different positions or measurement errors. This could also lessen the negative effect if the stimulation positions between the subject and *HeadSim 2* are not being identical. Also, having positions in the model with significantly different vibration components could possibly produce a more robust combination model. Thus, the positions chosen for the measurements

were far apart.

3.3.3 Threshold Measurements on a Test Subject

The subject from which the hearing threshold was obtained was the same subject on which *HeadSim 2* was based upon on. The test subject is a 39-year-old male with no previously recorded hearing loss. *HeadSim 2* is therefore assumed to be an anatomically correct representation of the skull, and the measured vibration data were matched with the hearing threshold of the subject. This is also because the anatomical structures affect the way sound propagates through the skull, and thus it is beneficial to perform the measurement on the subject which the model is based upon. While the literature is uncertain in regards to matching cochlear vibration with just one person's hearing threshold, an experiment setup such as this has not been tested before. Also, some of the literature suggests that the connection between an individual threshold and cochlear vibration could have a stronger connection when measuring all three axes, such as in this setup with *HeadSim 2*.

Stimulating the skull with a transducer generally gives responses from both cochleas. This is more prevalent when measuring hearing thresholds at stimulation positions such as the forehead where distance is about equal to both cochleas. Since the AC threshold was performed on one ear, and it was a desire to connect the AC threshold and BC threshold for a particular cochlea, masking was used. Normally, these types of thresholds are only performed up to 8000 Hz, and this is true for the masking noise as well. Because of this, a new masking calibration was performed for frequencies above 8000 Hz. Also, at higher frequencies ($>$ 10 000 Hz) the transducer tended to be audible in the air, which made it possible to hear through AC hearing since the subject did not have any conductive hearing loss. In order to make sure that the sound was heard through BC and not AC, *3M™E-A-Rsoft™FX Earplugs* was used for these higher frequencies, above 10 000 Hz, on the testing ear [37]. While it might have been possible to hear the transducer through AC at frequencies lower than 10 000 Hz, the ear was not plugged at these frequencies. The reason for this was to avoid the occlusion effect.

In order to limit the measurement errors present in audiometric tests, these tests were performed by a professional audiologist. Since certain frequencies were to be matched from each position, the tests were spread out over the period of two days; which were performed at the same time of the day. In order to ensure that the same position was simulated, photographs were taken. This meant around half of the frequencies were tested on both days, at all the different stimulation locations. The total time of conducted measurements of the subject was around 50-55 minutes for the first day and around 35-40 minutes for the second day.

3.3.4 Measurements on *HeadSim 2*

After measurements determining the bone conducting hearing threshold on the subject, the same vibration simulations were applied to *HeadSim 2*. The same actuator and metal band fixture was used in order to replicate the physical system involved in propagating the vibrations to the cochlea positions. Because of this, two layers

of the synthetic skin made from cured silicon, as mentioned in section 2.3, were used to replicate any dampening as realistically as possible. A setup with the *TestBand* and the synthetic skin can be seen in figure 3.4. The corresponding four stimulation positions were found on *HeadSim 2* and white noise was used to acquire the transfer function used to get the corresponding accelerometer readings for the bone-conducting threshold input voltage.

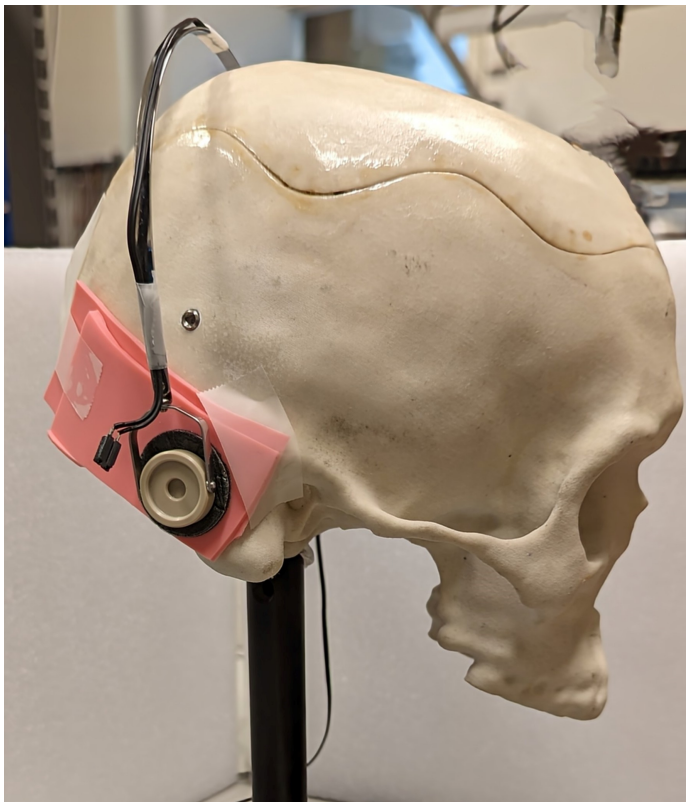


Figure 3.4: The test setup on *HeadSim2* at the mastoid position showing *TestBand* and two layers of the synthetic skin. A *BI300* implant can be seen at *Baha*® position.

While the audiometric tests with the *TestBand* are uncomfortable for the subject, the measurements on *HeadSim 2* could be conducted several times, with adjustment of the actuator position in-between test runs and then taking an average on the vibration data on each position. This was done five times for each position. The measurements were also performed for both the left and right accelerometers. This means when the measurements for the left accelerometer were performed, the actuator was moved to the left side on the mastoid and triangle position. For the frontal and back positions, they did not need to be moved and the measurement could simply be made on the same position but performing the measurements for the left accelerometer.

3.4 Implementation and Simulation

The proposed combination model was built with the data acquired from the mastoid, back, and frontal measurements. In order to assess the performance of the

combination model, the testing can be divided into two steps; offline and real-time simulation. Offline testing means analyzing the signals in MATLAB from the frequency response and real-time simulation means testing the signal processing chain in hardware.

3.4.1 Offline Simulation

Analyzing the performance of the combination model in an offline environment can be done in many ways. Here it was chosen to filter a 10-s audio file consisting of two people conversing with some minor background noise. This file consisted of an audio recording from a TV show, which means it was recorded with high sound quality, and thus it represented streaming the audio directly to the actuator. This audio signal was analyzed in both the time domain and the frequency domain. The chain of simulating the signal chain in the offline setting was implemented in MATLAB.

In order to simulate the accelerometer readings that would correspond to the audio file when streamed through an actuator attached to *HeadSim 2*, FIR filters corresponding to *HeadSim 2* had to be computed. However, since the frequency response functions used in system identification were already known, these could be used for this purpose as well. The measured frequency response functions were used to simulate the resulting accelerometer readings, one for each component $G_x(j\omega)$, $G_y(j\omega)$, and $G_z(j\omega)$. The position chosen to test the combination model was the triangle position since this position was not used in the determination of the model, and was thus deemed to be a good point of reference. As with the case of the frequency responses for the designed filters, these frequency responses corresponding to the triangle position had to span the full spectrum and be converted to the impulse responses to make it applicable to the time domain audio signal. The same interpolation method and assumptions outside of the measured frequencies were done as with the designed combination filters., i.e. PCHIP and an attenuating slope below 25 Hz and above the highest measured frequency value. The IFFT of the frequency responses was used to get the impulse responses in the corresponding FIR filters. The impulse responses were also shifted, truncated at 4000 taps, and windowed with a Hanning window. This was to ensure no wrap-around, and also make sure that every characteristic of *HeadSim 2* was captured. With this established, the audio signal was filtered through the FIR filters, resulting in three signals corresponding to the simulated accelerometer readings, which were then combined through the developed combination model which resulted in an audio output.

In the frequency domain, *Welch's power spectral density estimate* with a 50% overlap between segments was used of the audio input and output to represent the power distribution in the frequency domain. This visualized how the signal changed in the chain of filters and to see what happens to the signal and in which step. This was to draw conclusions regarding when the frequency component changed due to the actuator and *HeadSim 2*, or when it changed due to the proposed combination model.

The audio signal was also listened to before and after the filtering or the summation of vectors. However, listening to the signal can be limited in regard to quantitative

results and objectively assessing the quality of the result in this manner. Listening in an offline setting was instead used as a tool to give early insights into possible errors in the signal processing chain.

In order to further evaluate how the combination model is performing and how realistic it sounds, it was compared with the summation method proposed by Dobrev et al. By using offline simulation of an audio signal, algorithms that could be challenging to implement in a real-time scenario with low latency could be evaluated in this setting. This could be the case for the algorithm proposed in Dobrev et al. since it requires the DSP to efficiently handle trigonometry, logarithmic relations, loops, and if-statements. However, by evaluating it in an offline case where the signal chain is simulated, insights could be provided regardless of efficiency. A comparison of the audio output metrics, both in the time domain and frequency domain, between the combination model and the other method was therefore done.

3.4.2 Real-Time Simulation

In order to realize the combination model and perform the signal processing in real-time, implementation with low latency was required. A requirement of 10 ms was adopted in order to make the latency unnoticeable. This requirement put a limit on how computationally heavy any developed algorithms could be. It was also preferable to have the algorithm utilize the strengths of the DSP hardware. Because of this, the FIR filters created from the system identification were implemented on DSP hardware. However, since the structure of implementation for the method proposed by Dobrev et al. relied on other more challenging to implement efficiently in a DSP system for summation, as described in section 3.4.1, this was not implemented in the DSP hardware.

The noise affecting the whole real-time signal processing chain also needs to be minimized. Especially since the accelerometer output signal is low. Thus, keeping the noise floor of the DSP hardware to a minimum is of high importance to ensure a high SNR. Previous DSP hardware used for this purpose in *HeadSim 1* was *ADAU1701* from *Analog Devices* which measured an SNR of 100 dB for the ADCs and 104dB for the DACs [38]. Keeping the measured SNR above these values for any new hardware was of importance.

Further, following the layout of the signal chain, the system needs to process six analog inputs and output two analog outputs. However, since the two channels are processed independently the chain can be parallelized and utilize two pieces of processing hardware. This meant that only making sure that each channel connected to the corresponding channel input for the headphones was required. Thus, having two pieces of processing hardware with at least three analog inputs and one analog output is also acceptable where the x-, y-, and z-signal from the accelerometer are connected to the input of the DSP.

The hardware chosen for this objective was *ADAU1467 SigmaDSP audio processor* integrated on the evaluation board *EVAL-ADAU1467Z* [39]. The evaluation board made all of the functionalities of the chip available. An *AD1937 24-bit codec* was also integrated into the chip to handle the analog inputs and outputs. The codec had

four ADC with 107 dB SNR, and eight DAC with 112 dB SNR [40]. The *ADAU1467* DSP could efficiently handle FIR filters and be used for signal processing fulfilling the latency requirement using FIR filters with up to 24 000 taps at standard 48 kHz sampling time [41]. Two of these DSPs were used and each DSP implemented the signal processing for each channel. This DSP hardware was connected to the software *SigmaStudio©* where the filters could be realized and could be used for monitoring the signal processing.

The developed combination model was implemented in the chosen DSP via FIR filtering. In order to implement the FIR filters in *SigmaStudio©*, the FIR filter coefficients were desired to be less than 1. The gain of the filters for all frequencies was therefore adjusted so that this was fulfilled.

Other devices and connections are required to realize the real-time simulation setup so that it could be used for testing the sound quality. This is also how the system could be used in order to test the hardware of the components in the hearing aids. Firstly, a laptop was connected to an external *RME Fireface UC* sound card that streamed sound to the actuator through the built-in headphone amplifier. The actuator was connected to *HeadSim 2*, vibrating the skull accordingly. The six signal output from *HeadSim 2* was connected to two *EVAL-ADAU1467Z*, one DSP for each accelerometer in *HeadSim 2*. The output from the DSPs was then combined in a right and left channel and connected to a *Samson QH4* headphone amplifier in order to add volume control. Then a pair of headphones were connected to the amplifier. The setup used for real-time testing is seen in figure 3.5. This setup was used in order to verify that the signal processing chain was working as intended. It was also used to verify that the latency was regarded as non-noticeable by tapping *HeadSim 2*.

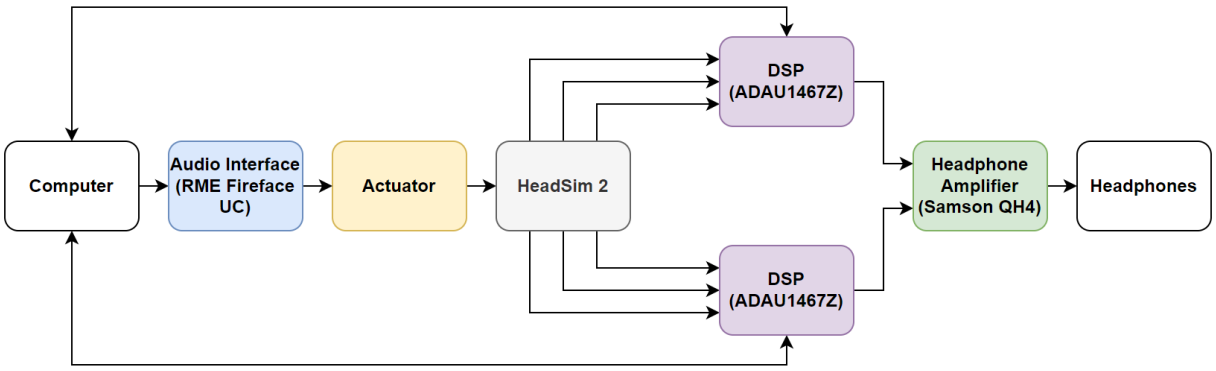


Figure 3.5: Schematic of the setup enabling testing the whole setup and model in a real-time environment.

The setup can be modified in order to change the vibration input to *HeadSim 2*. For example, the DSPs could run in a standalone operation mode when the computer, audio interface, and actuator are replaced by one or many bone-conduction hearing aids. In this way, the method of creating the vibrations in *HeadSim 2* can be

arbitrary and the rest of the setup will make the vibrations listenable by using the proposed combination model. One way of modifying the setup can be seen in figure 3.6. Here, a *Baha® 6 Max* hearing aid is used to induce vibrations into *HeadSim 2*. A reference microphone connected to a sound level meter is also used to analyze the impact of the *HeadSim 2* signal processing chain.

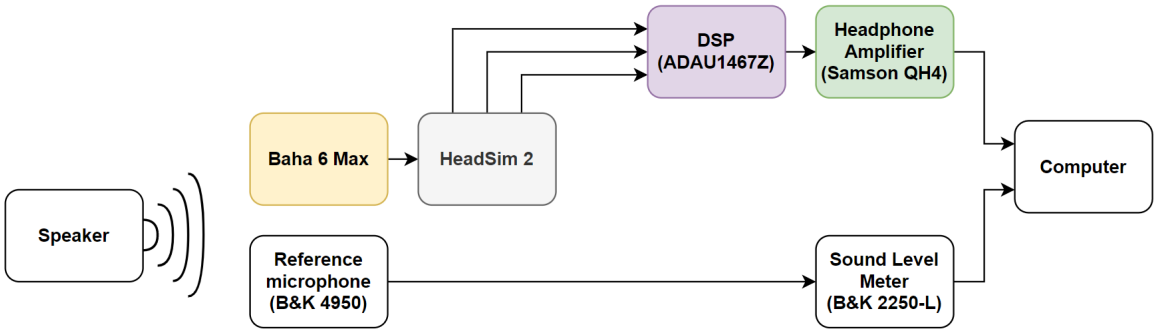


Figure 3.6: Schematic of the used test setup in a soundproofed room to compare the output of the combination model to a reference.

This setup was used in a soundproofed room where the reference microphone was situated close to the *Baha® 6 Max*. The audio was then played from a speaker and the power spectral density estimates were recorded and analyzed. The audio signal played from the speaker was a speech recording from a male voice in a quiet setting, with a similar sound quality as the sound file used in the offline simulation. The audio recording was played from the speakers with a volume of 65 dB SPL.

4
Results

The results are divided into four parts. The first part will present the results from the measurements; both the hearing thresholds made on the subject and the measurements made on *HeadSim 2*. The second part will be dedicated to the resulting FIR filters derived from the data with the created method. In the third part, the FIR filter performance will be analyzed offline, in MATLAB, and be compared against the method proposed by Dobrev et al. The last part will present the result from the real-time simulation, which means filtering live audio with the DSP hardware.

4.1 Threshold and Vibration Measurements

In this section, the measurements from the audiometric test of the subject, the measurements of *HeadSim 2*, and a comparison of the data from both of these measurements, will be presented.

4.1.1 Hearing Thresholds of Test Subject

The measured AC threshold for the right cochlea can be seen in figure 4.1, where a normal hearing curve can be observed with the exception of 14 000 Hz, where the threshold is significantly higher compared to the rest of the frequencies.

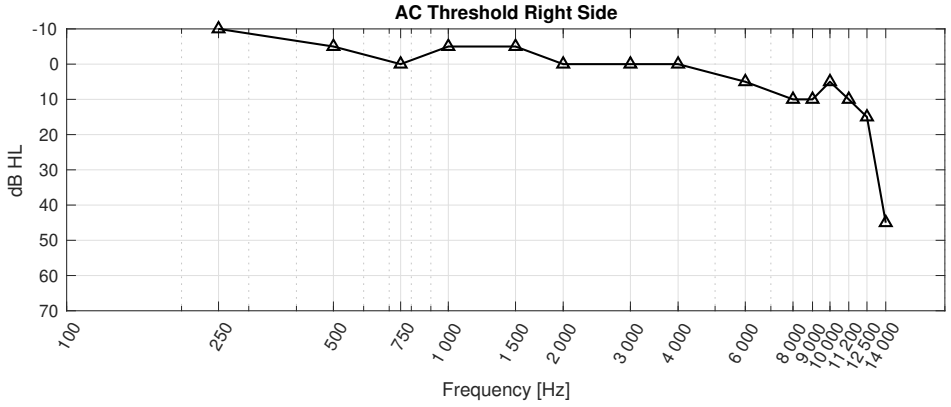


Figure 4.1: AC threshold of the right cochlea from the subject.

The mastoid position BC threshold seen in figure 4.2 follows the AC threshold but with a slightly higher threshold overall. The rest of the positions follow a similar pattern as the mastoid position, with the exception of the lower frequencies and the 14 000 Hz mark, where the mastoid has lower threshold values.

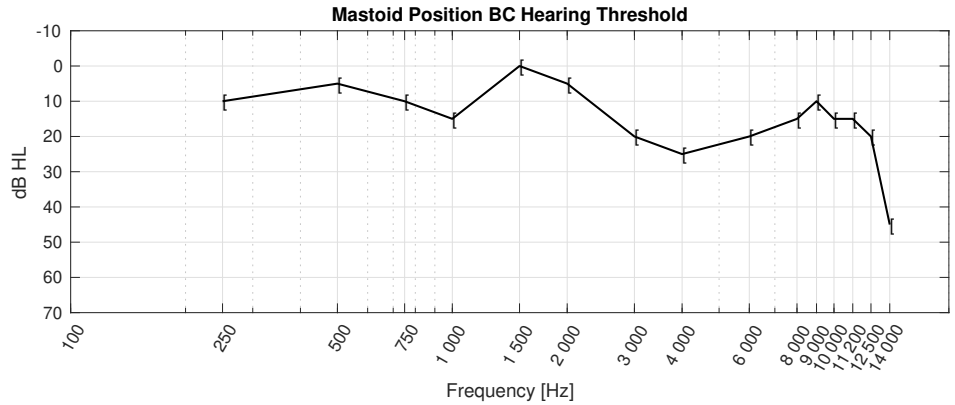


Figure 4.2: BC threshold of the right cochlea stimulating at the mastoid position.

The triangle position as seen in figure 4.3 yielded the most similar result to the mastoid position, and in some places such as 1000 Hz it produces a slightly better threshold.

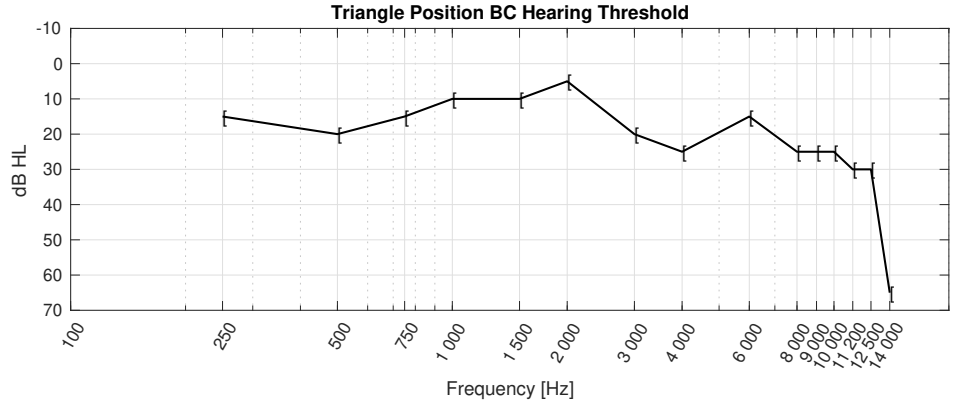


Figure 4.3: BC threshold of the right cochlea stimulating at the triangle position.

The back position seen in figure 4.4 has a higher threshold in general compared to both the mastoid and triangle positions.

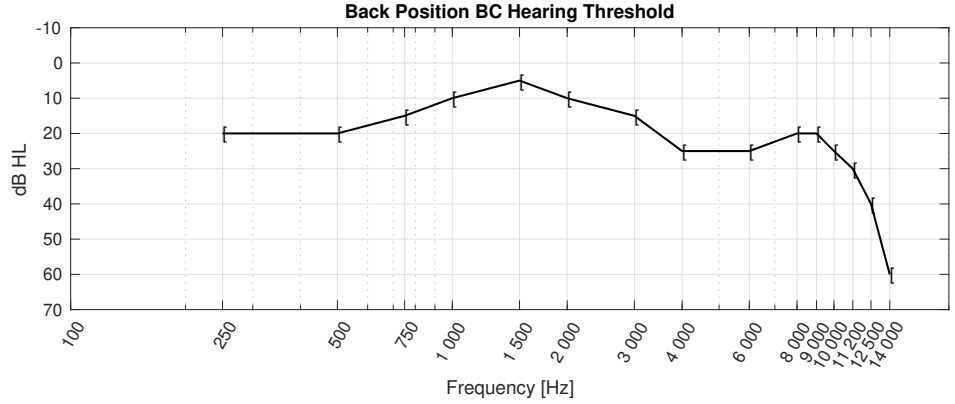


Figure 4.4: BC threshold of the right cochlea stimulating at the back position.

The frontal position seen in figure 4.5 has the highest threshold but overall resembles the threshold of the back position.

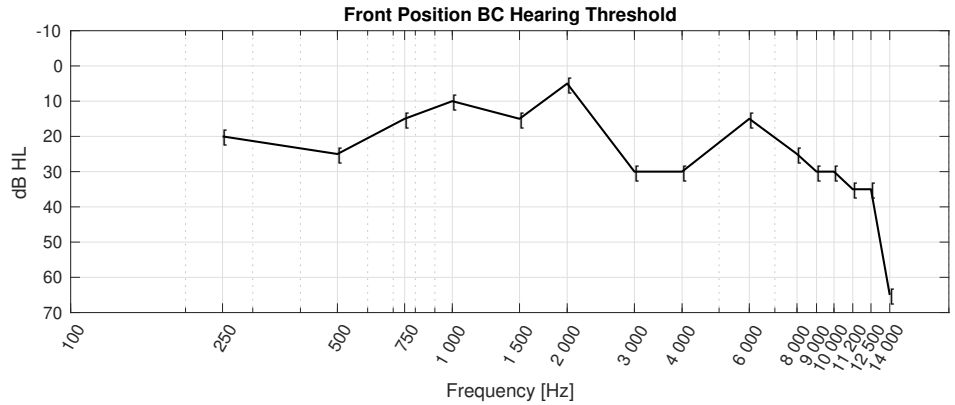
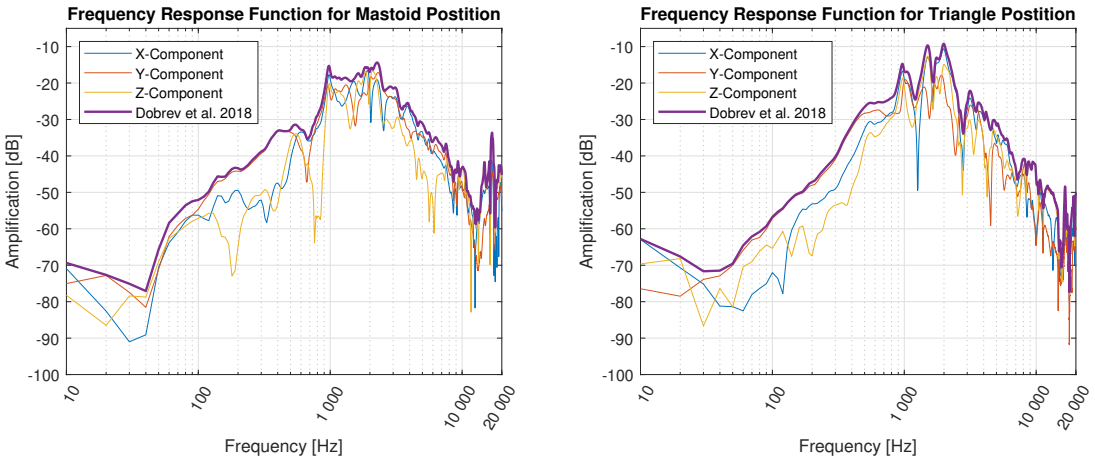


Figure 4.5: BC threshold of the right cochlea stimulating at the frontal position.

4.1.2 Measurments on *HeadSim 2*

In the figures below, the resulting frequency response function (FRF) of *HeadSim 2* can be seen. Because the accelerometer uses 3 axes, and the hearing threshold is one-dimensional, a combination of the components using the method described by Dobrev et al. is also plotted. This is done for the sake of later comparisons since it is also noted by the authors that this is more closely related to the BC hearing experience. Since both the left and right accelerometers produced similar frequency responses only the frequency responses of the right side are shown in the figures below.



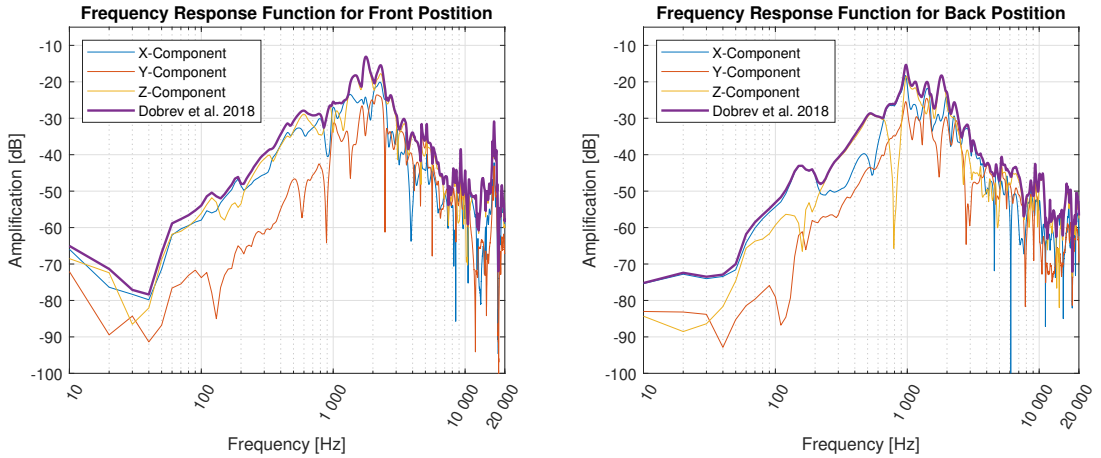
(a) FRF at the mastoid position.

(b) FRF at the triangle position.

Figure 4.6: FRFs in *HeadSim 2* stimulating at the mastoid and triangle position with two layers of synthetic skin.

Studying the FRF for the mastoid position in figure 4.6a and triangle position in figure 4.6b, the resulting amplification of the combined data is similar. In some areas,

the mastoid position yields a higher amplification, such as in the lower frequency areas between 250-600 Hz, and the triangle position yields higher amplification between around 600-1000 Hz. In terms of the individual components, both positions see the y-component being dominant in the lower frequency regions, and all components being quite similar at frequencies above 1500 Hz, except for the z-component being slightly lower at the mastoid position. It can also be noted that both the x- and z-direction are more dominant, especially the x-direction, in the triangle positions at lower frequencies compared to the mastoid position.



(a) FRF at the frontal position. **(b)** FRF at the back position.

Figure 4.7: FRFs in *HeadSim 2* stimulating at the frontal and back position with two layers of synthetic skin.

The FRF for the back and frontal position seen in figure 4.7b and 4.7a respectively produces higher attenuation compared to both the mastoid and triangle position. The back and frontal position shows similar behavior in both the combined curve and the individual components. Compared to the mastoid position, both of these positions exhibit much lower values for the y-component, and a mix between the x- and z-components is dominant depending on the frequency and position. Due to the rigid body dynamics, this is more noticeable for the lower frequencies. Moreover, since the axis separation for the accelerometers is around 24 dB, the true differences can be more than shown in the figures when the differences between the axis are close to this value.

4.1.3 Comparison of Data

By comparing threshold differences and frequency responses between the measured positions, insights about what can be a representative indicator for BC hearing in *HeadSim 2* can be given. This comparison is seen in figure 4.8. The figure compares the y-component and the summation from Dobrev et al. to the measured hearing threshold differences. Comparing the FRF and hearing threshold for the triangle position to the ones at the mastoid position shows a discrepancy at the lower frequencies where neither of the y-component or the summation follows the

differences in BC hearing at these frequencies. This can suggest a poor choice of BC hearing indicators, or that the measurements at these frequencies contain errors. However, as shown when comparing the triangle and frontal position, the summation from Dobrev et al. does only deviate from the threshold differences with maximum around 10 dB, compared to the y-component which fluctuates and sometimes differ with significantly more than 20 dB.

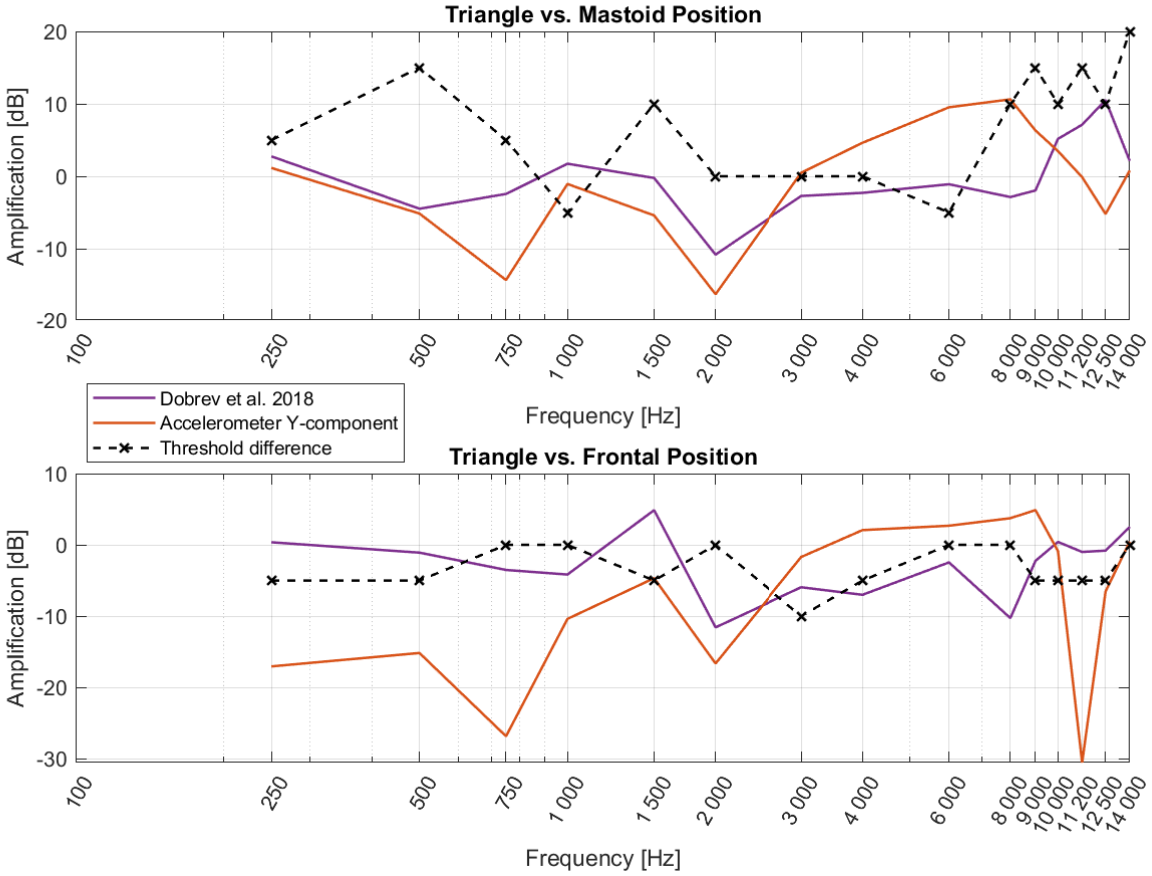


Figure 4.8: Comparison between differences in hearing thresholds and FRFs between positions. The currently used BC hearing indicator, the y-component, and a method of combining the spatial components from Dobrev et al. are shown.

4.2 Model

In this section, the resulting filters based on the proposed combination model created based on the hearing threshold and measurements on *HeadSim 2* will be presented. Since there was a large increase in the hearing threshold at 14 000 Hz and combined with the unreliable behavior of warble tones at higher frequencies, this data point was excluded from the model. Also, the data for the mastoid position was modified in the first five frequency points into the same threshold values as for the triangle position. Without this modification, the filters had a tendency to have dubious behavior at lower frequencies, and comparing all the different positions' thresholds

at the lower frequencies indicated that some acoustic transmission had taken place at the mastoid position.

4.2.1 Designed Filters

The designed filters can be observed in figure 4.9. It can be noted that frequencies outside of the interval used in the system equation determination have been assigned a value. This means that the filters have the same gain and phase angle between 250 Hz and 25 Hz as the respective filters have at 250 Hz. Below 25 Hz and above 12 500 Hz, a -20 dB per decade slope has been added.

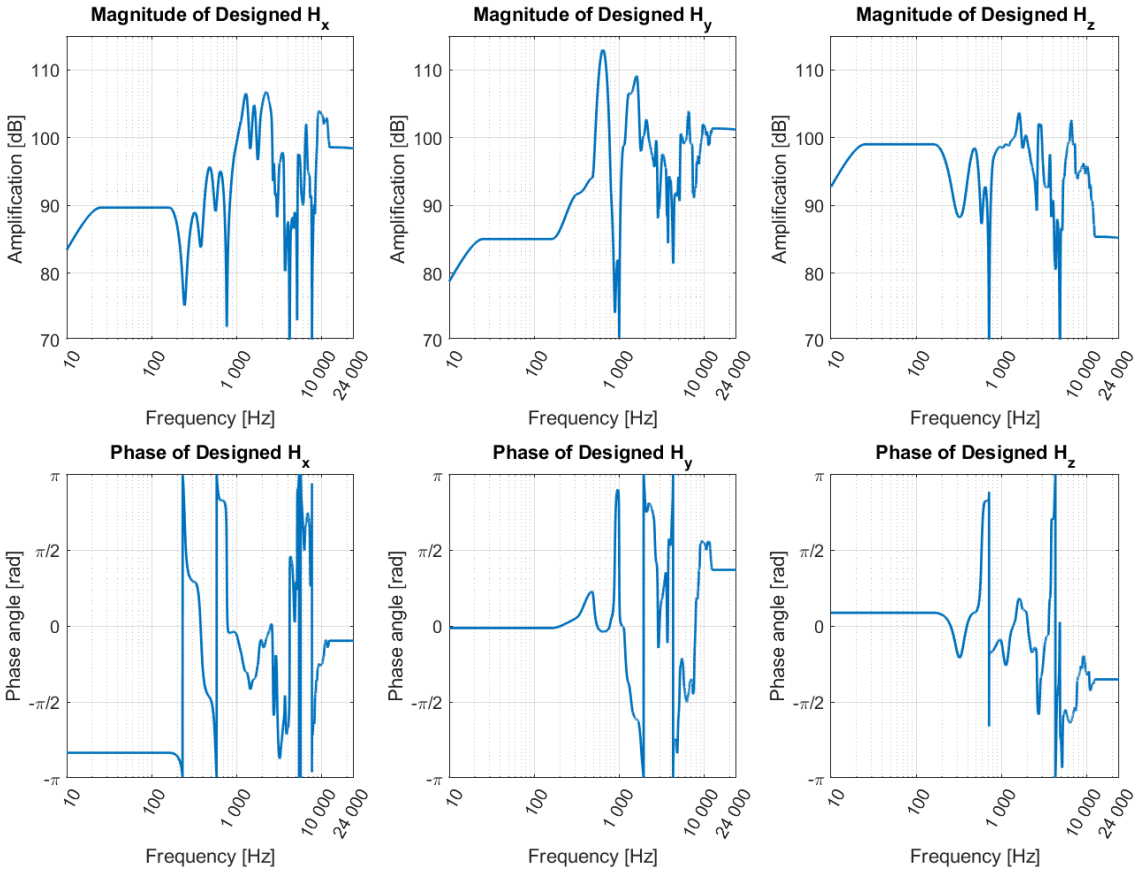


Figure 4.9: The transfer functions of the FIR filters used to implement the combination model for the right side. The designed transfer functions span the whole spectrum.

4.2.2 Filter Output

The output gain from the filters can be seen in figure 4.10 and the output phase in figure 4.11. It can be seen that the filter output has a better fit for the positions used in the determination, i.e. mastoid, back, and frontal positions, compared to the validation position, i.e. triangle position, especially at frequencies below 1000 Hz. It should be noted that without the sinc interpolation and third-octave smoothing, the

positions that the model is based upon would have had the same estimated output as the original output. Without those implementations, however, the match for the validation position would have been worse.

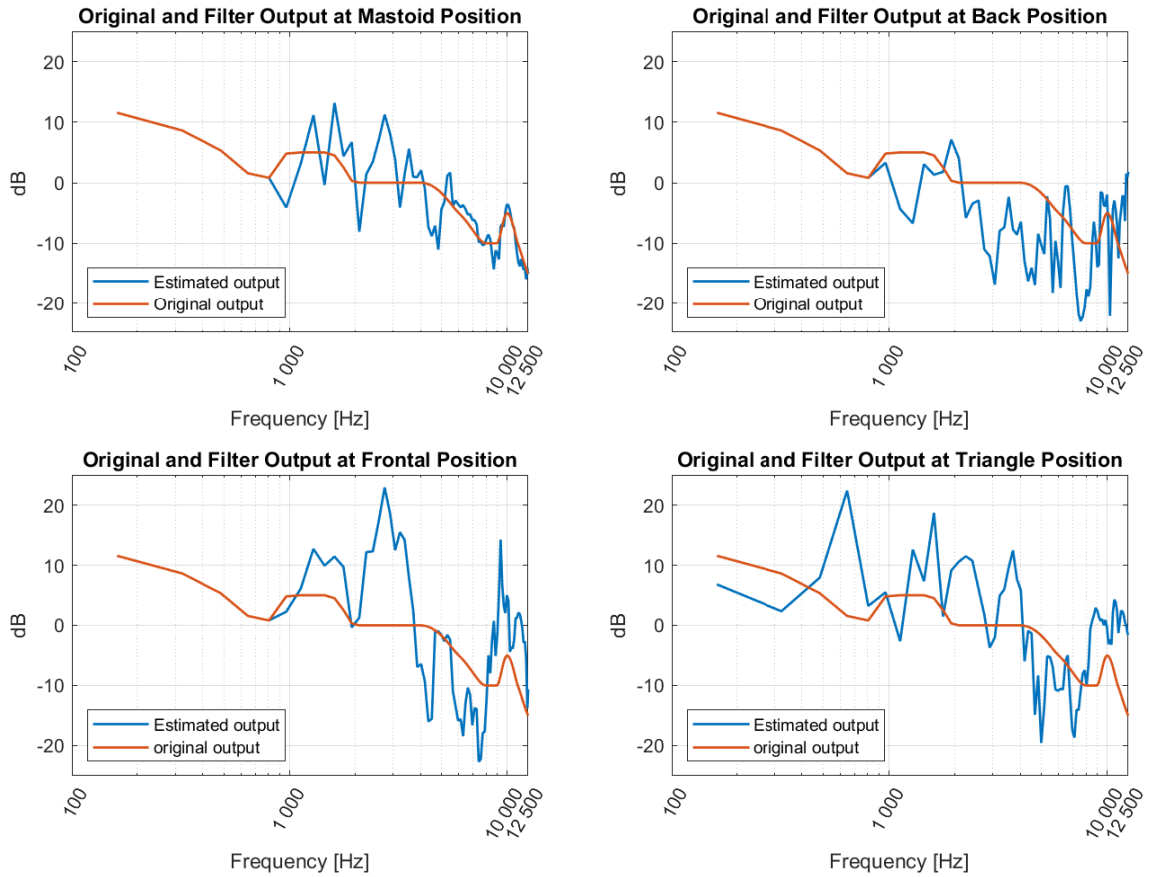


Figure 4.10: The measured magnitude of output compared to the estimated magnitude of the output corresponding to the right combination model.

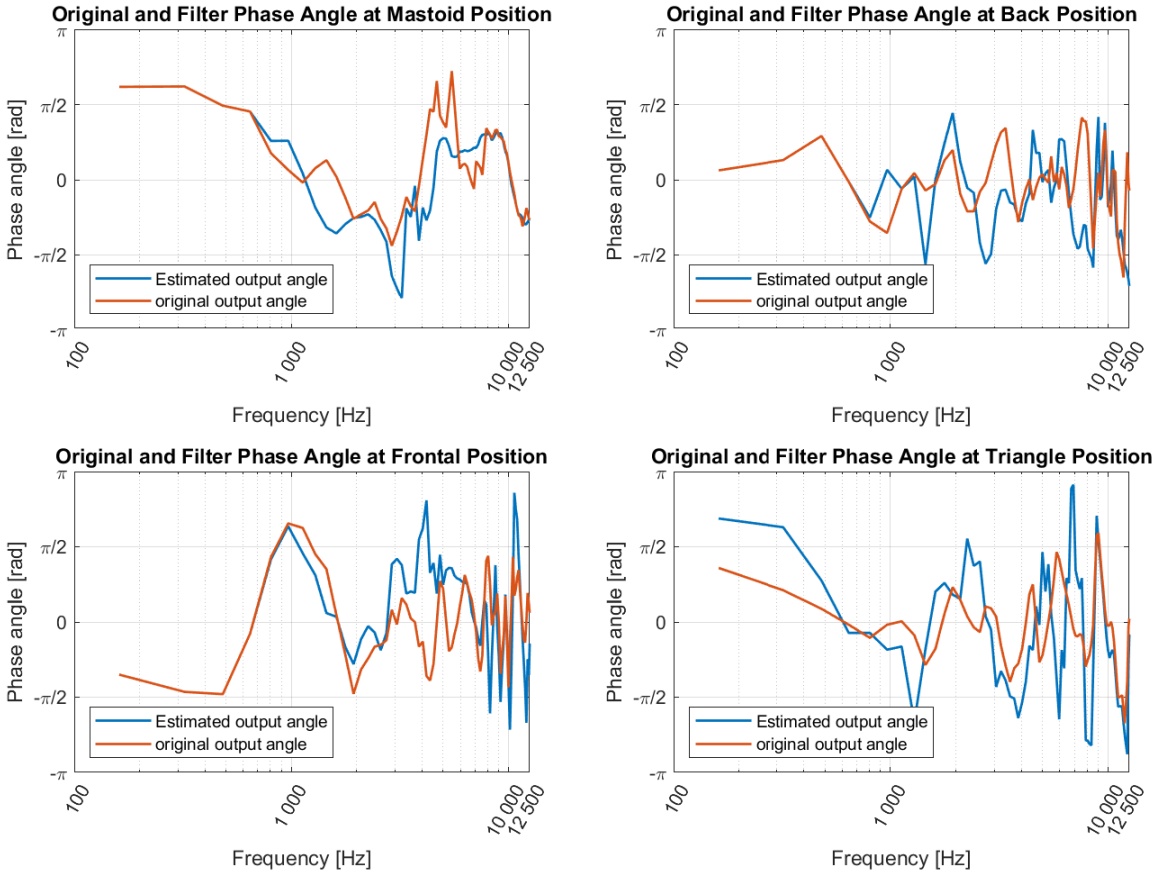


Figure 4.11: The measured phase angle of output compared to the estimated phase angle of the output corresponding to the right combination model.

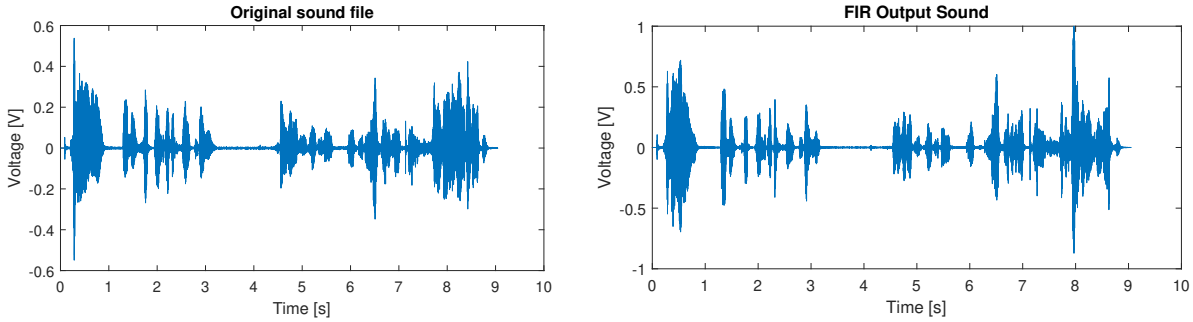
4.3 Offline Testing

To test the resulting filters in an offline setting, there were several tests were performed. To judge the performance of the filter, the triangle position was chosen since this was not part of the system determination. The output was then analyzed in both the time and frequency domains with the help of the 10-s sound file. The combination model was also compared to the method of adding the components together proposed by Dobrev et al. at the triangle position. In this section, only the right position is shown, since it was seen that both sides produced a similar result and that there was no effort to listen to stereo sound in offline testing.

4.3.1 Sound File Simulation

The 10-s sound file, which can be seen in the time domain in figure 4.12a, was processed through the skull system and filtered through the designed FIR filters, which could then be compared with the original sound file. The filtered signal output in the time domain can be seen in figure 4.12b. Comparing the two signals,

there are some noticeable differences, such as a higher peak at the 8-second mark of the filtered output.



(a) 10-s audio signal input to the offline simulation.

(b) 10-s audio signal output from the offline simulation.

Figure 4.12: The sound file in time domain before and after the filtration chain simulating the system of *Head Sim 2* and the combination model.

The frequency components of the signal can be visualized in the frequency domain. The power spectral density estimate of the individual accelerometer components after simulation of the head simulator can be seen in figure 4.13. In the figure, the power of the spectrum is shown and is rather similar across the different axes. However, it is noted that the y-component has a slight increase in power for lower frequencies when the triangle position is used.

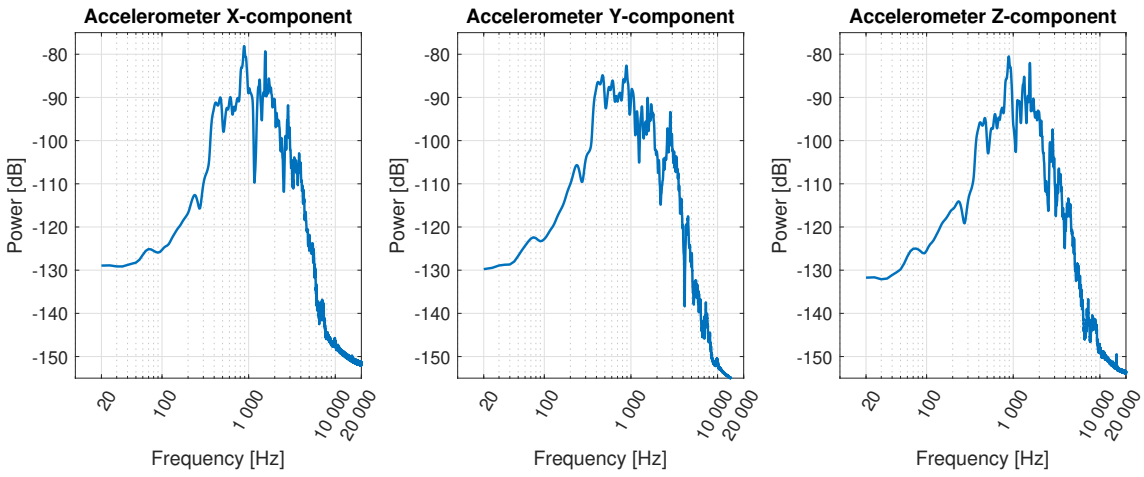


Figure 4.13: Power spectral density estimate using Welch's method after simulation of the accelerometer readings from *HeadSim 2* when inputting the sound file at the triangle position.

Using the simulated accelerometer data when inputting the sound file, the combination model can be tested and the spectral component can be visualized. The spectral density estimate of the outputted audio signal can be seen in figure 4.14, where both the original sound file and filter output can be observed. Both the original sound file and the outputted audio signal have been adjusted to keep the time domain signal

in the interval of $[-1, 1]$ to make the signals listenable in MATLAB. It is clear that the lower frequency components are attenuated greatly, which can originate from the actuator and *HeadSim 2* system as this is also seen in figure 4.13. Overall, the output from the combination model combines the accelerometer readings, and the frequency components with the most power are between 300 and 4 000 Hz.

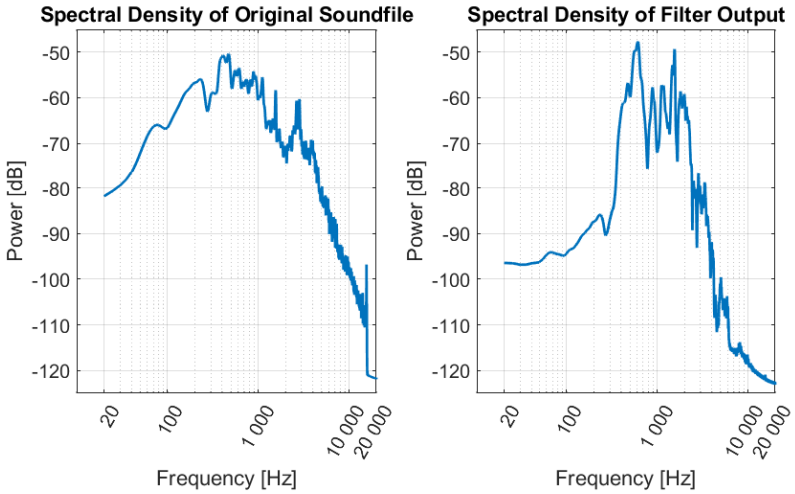


Figure 4.14: Power spectral density estimate using Welch's method after simulation of the accelerometer readings and the combination model when inputting the sound file at the triangle position. Both the original inputted sound file and audio signal output are adjusted to keep the amplitude of the time domain signal in the interval of $[-1, 1]$.

By listening to the audio signal output, it is clear that the lower end of the frequency spectrum is attenuated. The bass in the voices of some of the speakers is not as present in the simulated output, compared to the original sound file. It can also be noted that the simulated sound has a metallic sound quality. Additionally, the rustling of paper and other higher-frequency components of speech are more dominant in the simulated output. Comparison between combining the simulated accelerometer data in the proposed combination model and through the method in Dobrev et al. only show minor differences when listening to the output in headphones. A slight difference can sometimes be noticed where some parts, i.e. frequency components, are more or less noticeable.

4.3.2 Comparison with Other Methods

In figure 4.15, the frequency response of the x-,y-, and z-components can be seen together with the frequency response of the designed combination model and the method described by Dobrev et al. These two methods add the three components together in different ways, but it can be observed that they follow similar behavior across the frequency range with the exception of some frequencies. A moving warble band mean has been applied in the visualization of the output from the proposed combination model to include the $\pm 5\%$ warble band around the center frequency. A

relatively large amplification can be seen at 600 Hz and a relatively low amplification can be seen just before 5 000 Hz.

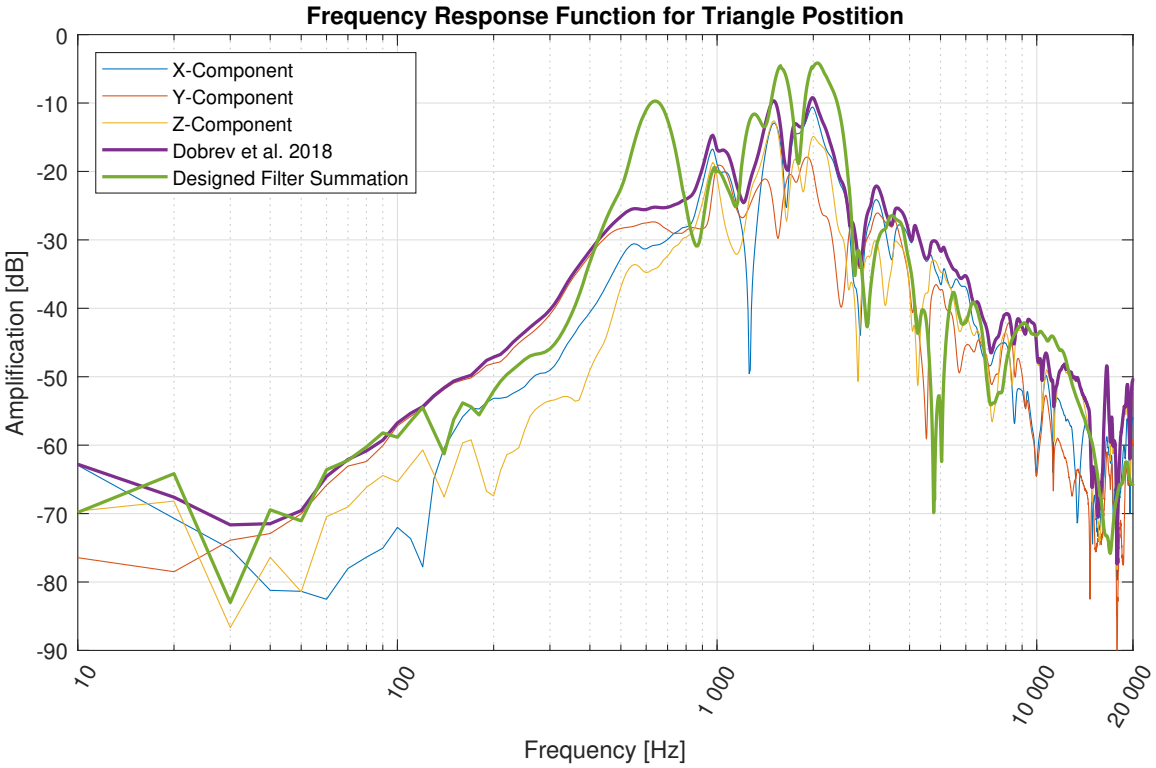


Figure 4.15: Frequency response function for *HeadSim 2* at the triangle position. The individual orthogonal accelerometer components are combined in either the combination model through the designed filter summation or the method from Dobrev et al.

Comparing the differences in the summation methods to the threshold differences can once again give insights into what might be a representative indicator for BC hearing. In figure 4.16, it can be seen that the adjusted threshold for the mastoid position at the triangle and mastoid position are identical for most of the data points. The combination model output through the designed filter summation is visualized with a moving warble band mean to display the differences in this frequency band. It is also shown that the combination model is performing similarly to the method from Dobrev et al. at following the threshold differences between positions while mitigating the sharp and big deviations followed by solely observing the y-component.

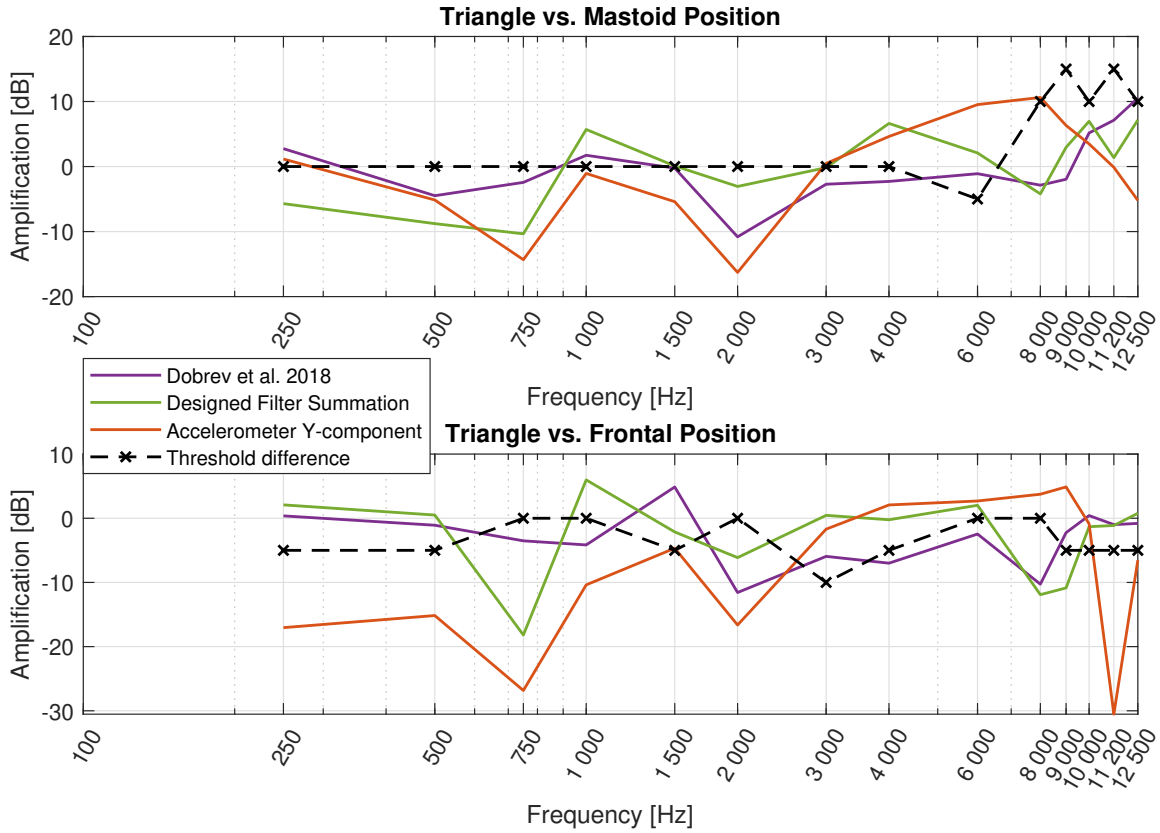


Figure 4.16: Comparison between differences in hearing thresholds and FRFs between positions. The currently used BC hearing indicator, the y-component, a method of combining the spatial components from Dobrev et al., and the proposed combination model are shown.

When listening to the simulation of the combination model in an offline environment, there is no notable difference between the left and right accelerometers corresponding to this model.

4.4 Real-Time Testing

By making use of DSP hardware and efficiently implementable algorithms the combination model was implemented through FIR filter summation with the FIR filter coefficients for the right accelerometer seen in figure 4.17. The FIR filter coefficients for the combination model for the left accelerometer are structured in a similar way. The coefficients with the greatest magnitude are centered and it can be noted that 1500 taps are enough to make the filter coefficient dampen at the ends, mitigating the impact of the window function.

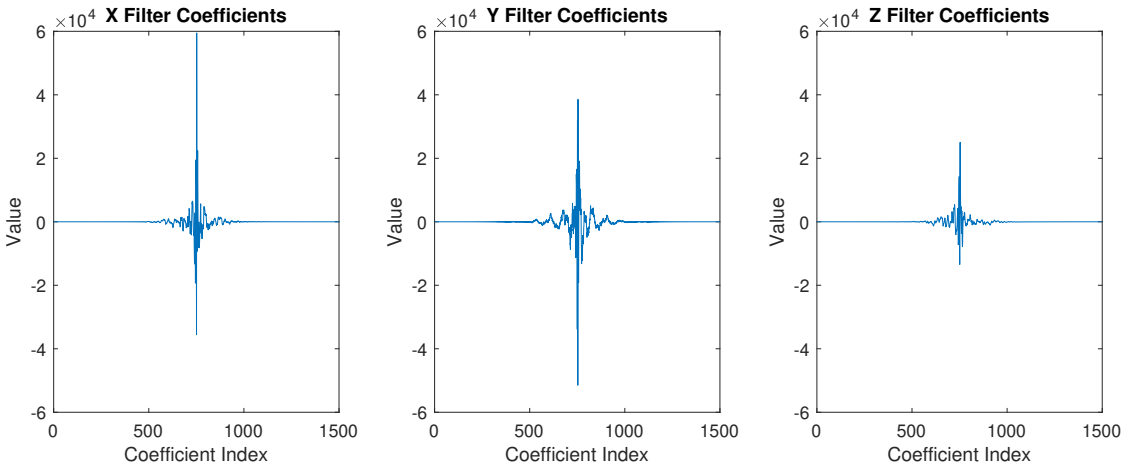


Figure 4.17: The impulse responses for the FIR filters implementing the combination model for the right accelerometer in the DSP hardware.

The transfer functions for the FIR filters implementing the combination model for the right accelerometer can be seen in figure 4.18. It can be noted that the appearance is closely following the designed filters, with some changes in the lower frequency components.

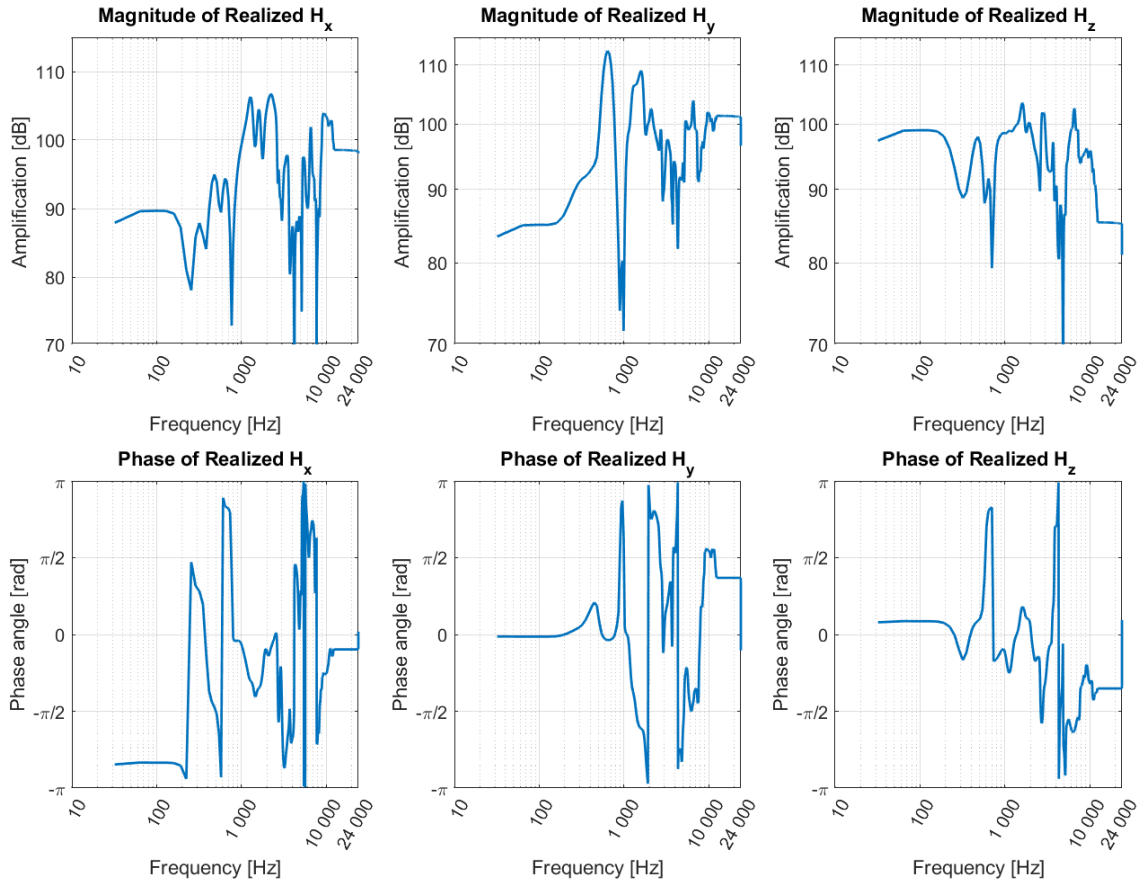


Figure 4.18: The transfer functions for the FIR filters implementing the combination model for the right accelerometer.

The transfer functions for the FIR filters implementing the combination model for the left accelerometer can be seen in figure 4.19. Some general resemblances can be found when comparing the transfer functions of the realized filter for the right and left sides. However, the data used for determination results in different designed filters for each side, as seen in the figure.

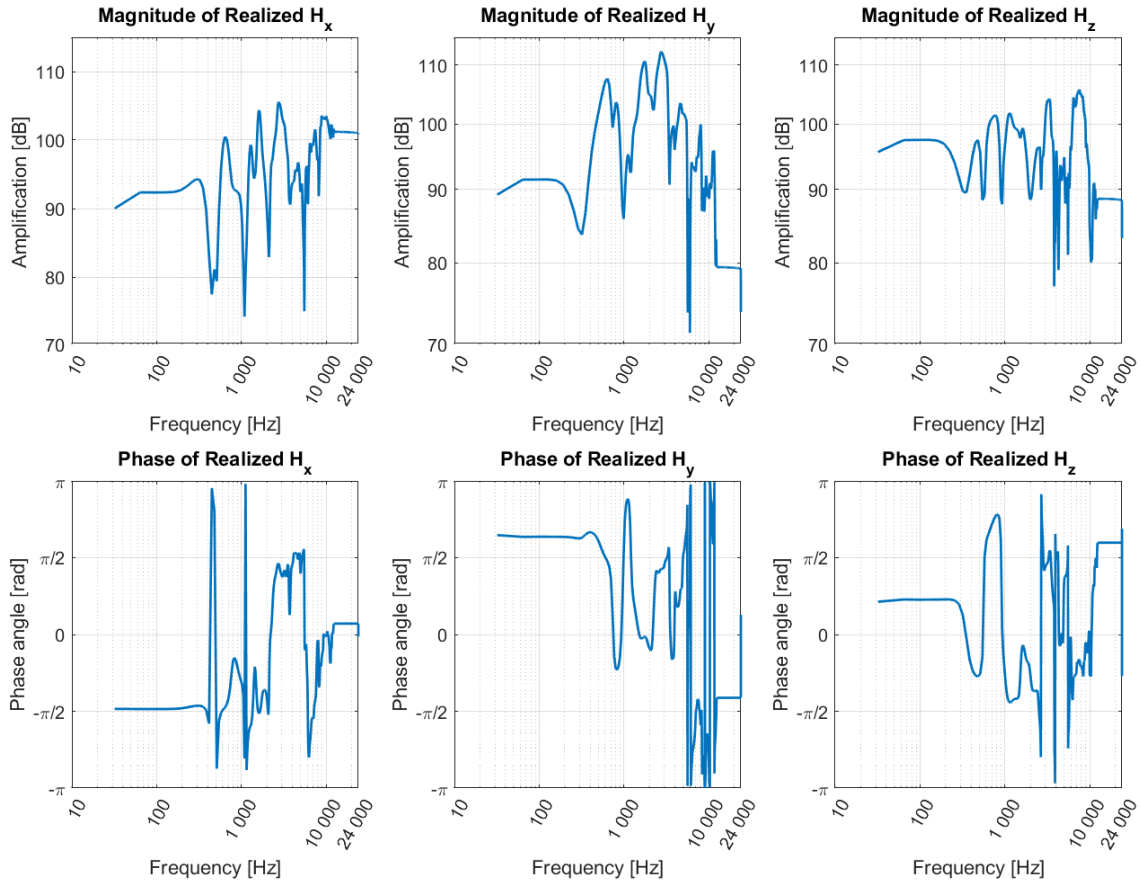


Figure 4.19: The transfer functions for the FIR filters implementing the designed combination model for the left accelerometer.

It can also be noted that similar to the offline testing, the models for the left and right accelerometers have similar sound quality. However, when stimulating in real-time, the filter gain of the two combination model sides is slightly different due to the adjustment made to keep the FIR filter coefficients below 1.

Playing an audio recording of a male voice in a soundproofed room with a commercialized BC hearing aid shows that the signal chain properly outputs an audible signal. A comparison between the reference microphone recording and the combination model output in terms of power spectral density estimate can be seen in figure 4.20. It can be noted that it shows similarities between the simulated result in 4.14 when it comes to changing the spectrum of the reference signal. However, the real-time simulation also shows a greater ripple, with fluctuating power, over 600 Hz. Also, a relatively big increase in power can be seen for higher frequencies. This shows a tendency of amplifying some higher frequency components to a larger extent compared to the reference microphone, possibly giving a high amplification to overtones originating from the voice.

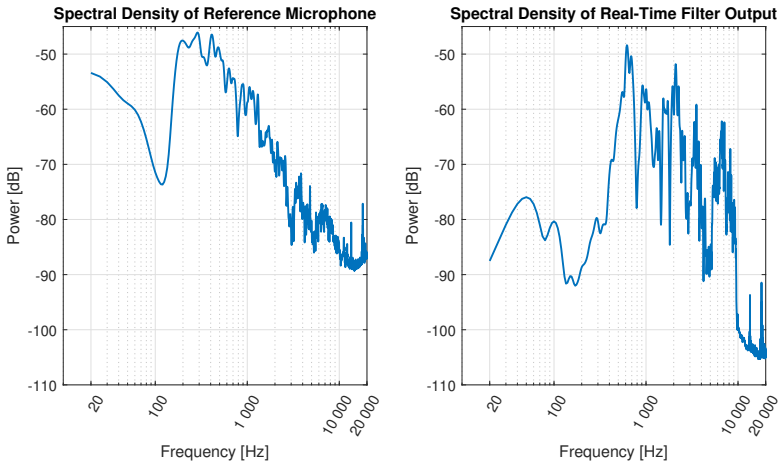


Figure 4.20: Power spectral density estimate using Welch's method of the reference microphone signal and the combination model when playing a recording of a male voice in a soundproofed room. Both the audio signals are adjusted to keep the amplitude of the time domain in the interval of $[-1, 1]$.

5
Discussion

In this part, the results will be evaluated and further commented on in order to discuss causes and outcomes. It is divided into four main parts, where all sections from the result are included, in addition, future work is discussed within each section.

5.1 Measurements

The proposed model which relies on system identification is greatly affected by the quality of the data in various ways. This is not only true in regards to how the data was acquired but also to the amount of data points chosen in terms of stimulation positions and frequencies tested.

5.1.1 Modification of Data

Since the quality of the data was very important, it was necessary to make some adjustments to the acquired data before implementing the method of assembling the combination model. This includes removing the 14 000 Hz measurement point and changing the mastoid position hearing threshold to have the same values as the triangle position for lower frequencies.

The 14 000 Hz data point was deemed to lie outside of the hearing limits of the subject since it was a relatively big increase in the threshold at this point. Combined with the fact that warble tones were being used at these higher frequencies suggested that these data points impacted the model with uncertainty. Keeping this frequency point made the model overestimate the gain of the filters at higher frequencies, which amplified these frequencies and affected the system output quality in a negative manner.

Before modifying the mastoid hearing threshold, it was noted that the model behaved in a way that was unrealistic in the frequency range of 250-1000 Hz, which consisted of four measurement points. Studying both the hearing threshold and the FRF measurement on *HeadSim 2*, a miss-match at the mastoid position at this frequency range could be seen. This could come from either faulty data from either the threshold or the *HeadSim 2* measurement. While it could indeed be something faulty with the *HeadSim 2* measurement, such as with the measuring procedure itself or relating to the construction of the skull, it was deemed more likely to be an issue with the audiometric measurement. This is partly due to it only being an issue at a specific frequency range at a specific stimulation position; if there were major issues *HeadSim 2* it would probably show up at other places as well. The test on *HeadSim*

2 was also repeated with similar results. However, the major reason the issues were attributed to the threshold is that acoustic transmission is a known issue at lower frequencies, and thus it was assumed to be the underlying issue. At this stage, it was decided that the closest position in both vibration data and the location was the triangle position, and thus the first four frequencies seemed to provide a better fit for the mastoid position as well.

5.1.2 The Effect of Stimulation Positions

The chosen positions probably affect the model quality a lot. Even though the mastoid position had dubious behaviour at lower frequencies, the performance of the filters was affected in an even more negative manner if the mastoid position was discarded in favour of one of the other positions. For example, the model could have been built from the triangle, back, and frontal positions instead, but this made the filter behaviour a bit worse in terms of how busy the output would be. As described in the method, the positions were chosen to be as far away from each other as possible, in order to lessen the effect of any measurement errors.

It is also worth noting that the positions also had to be realistically chosen in terms of that they actually had to be tested on a real subject, which limits the amounts of positions actually possible. For example, the top of the head could be an interesting position to stimulate since it is located from a completely different angle, i.e. parallel to the z-direction, compared to the positions actually chosen, but this was deemed very difficult to perform on the subject with the *TestBand*.

5.1.3 Measured Frequency Data Points

The idea with the stimulation frequencies was to use as wide of a frequency range and as many frequency points as possible, without having the testing time for the subject being too long. The frequency range is limited by the subjects hearing threshold, which should be noted when the AC threshold is performed. It is not entirely clear how beneficial it is to stimulate at these higher frequencies above the normal 8000 Hz threshold, however, with the method chosen the frequency range should probably span the subjects threshold in order to capture the proper hearing behaviour. Also, testing for more frequencies below 250 Hz is perhaps even more relevant, since this range covers more octaves and is probably important for the perceived hearing. However, this is limited by the actuator used, which generally distorts the sound when the frequency is too low. Thus, it is perhaps more relevant to investigate how to realistically model the behaviour of this frequency range without any data point, and improve on the existing method of extrapolating these data points.

In terms of the number of frequency points, this is also something that can be tested further. The way this implementation works is by interpolating data points in between the tested frequency points. This was assumed to be reasonable since the thresholds tend to not fluctuate greatly between measurement points. However, this could be tested further. On the other hand, this could mean that fewer frequency points could be used to accomplish a similar performance of the filter. Studying the threshold for the mastoid position in figure 4.2 however, it is rarely completely flat,

which indicates that some information would be lost if fewer data points had been acquired. The same can generally be said for all of the thresholds. To assess more closely how this can affect the resulting filters and what gives the best performance, more testing would need to be conducted.

5.1.4 Warble Tones Effect on the Model and Data Acquisition

While warble tones were used during the audiometric tests, the way the data was acquired during the stimulation on *HeadSim 2* was different since it used white noise. It is not entirely clear at what stage this affected the matching of these two measurements. One idea with the usage of warble tones as opposed to stable tones was that a frequency band was tested instead. This is assumed to avoid any notches that could possibly be present, which can be seen in the accelerometer data in figure 4.7; if a frequency point is located at a notch it was deemed better to test a range surrounding this notch. This could be the reason why an improvement is seen when smoothing the FIR filters which implemented the combination model.

Testing a frequency band could in theory also affect this matching in a negative way as well, since it is not exactly clear which of the frequency components within the band that the subject actually hears. This was apparent at the higher frequencies where the 14 000 Hz data point was excluded. This frequency was probably outside of the subjects threshold but it was slightly audible anyways, which indicated that the lower end of that frequency band could be heard. This behavior could also be present at other frequencies as well, which could have led to an overestimation of the hearing threshold at some of the frequencies, probably most notable at higher frequencies.

It should be noted however, with or without warble tones, if a tone is audible but has a very high threshold value, it might be necessary to exclude such a point when creating the model even if it technically was audible for the subject. The reason for this is that the model assumes that the subject has normal hearing, and a too-high threshold of a certain frequency is in opposition to this assumption. This could make the model overestimate attenuation at this particular frequency which then makes for a poor result. It is not exactly clear how high a threshold needs to be for it to be considered poor enough hearing, but the data containing the 14 000 Hz point did partly invoke this behavior of the model.

5.1.5 The Effect of Differences Between *HeadSim 2* and the Test Subject

The way of acquiring the FIR filters does not necessarily depend on that *HeadSim 2* and the subject are identical; the model should take into consideration any absolute differences between the two. However, the more different the two are, the less general the model will be. While it was never intended to make a general model per se, it is something that could be interesting for future work/research in the area. Problems that may arise from having a non-general model are that the model might perform

differently if deviations between different head simulators exist. Differences can for example come from material choices and the accelerometers. The materials chosen for *HeadSim 2* do in general mimic a real skull well when studying certain properties such as proper attenuation, but they do not completely match the true materials. As for the accelerometers, they do not exactly match the cochlea in terms of weight, appearance, or location in the skull. One intuitive example is if the accelerometers in *HeadSim 2* are slightly misaligned, the developed combination model is then more unique to that head simulator. This might not exactly be the case but could apply to other aspects, making the model less general. It is difficult to assess how much they differ from each other in order to assess how this model would perform on another head simulator, and it could be that the measurements that the model is based on need to be performed each time a new head simulator is used.

Another aspect is related to if it is reasonable to only measure one accelerometer at a time in *HeadSim 2*, while the measurements conducted on the test subject will have a hearing response on both cochleae for an arbitrary vibration input. This is especially relevant at stimulation positions where the distance between the two cochleae is more or less equal, such as the back and frontal positions. To make up for this, masking was implemented with the goal of canceling out the non-tested cochlea. This is probably a simplification in some regards.

However, some deviations between the head simulator and the subject can affect the resulting filters even if it is not desired to make a general model. This is in relation to any relative differences between stimulation positions to due a difference in wave transmission for higher frequencies, where rigid body dynamics is not dominant. A large difference between material between *HeadSim 2* and the head of the subject could mean that the vibration difference between two positions does not correlate in the same way, and thus the connection to a hearing threshold becomes uncertain. For example, a real skull has sutures, which affect wave transmission, which are not present in *HeadSim 2*. However, judging from the results, it does seem that making the assumption that they are similar enough provides a reasonable result judging from the threshold and vibration response changes, but it should be kept in mind that this is somewhat of a simplification.

Another way of confirming that *HeadSim 2* and the test subject are similar enough is by studying the relative difference in threshold as seen in figure 4.8. Here it can be seen that the threshold data of the subject follows changes in the measured data of *HeadSim 2* when all axes are taken into consideration. This should then indicate that the quality of *HeadSim 2* is rather good and that the aforementioned possible issues with the said head simulator are probably not affecting the data enough to be a problem. This also indicates that some simplifications made, such as assuming that masking perfectly isolates the measured cochlea and connecting it to one accelerometer, are reasonable assumptions.

5.1.6 Comparison of Indicators for BC Hearing in *HeadSim 2*

Comparison between how the BC hearing thresholds change compared to how the frequency response from *HeadSim 2* changes can provide insight into what might be a suitable descriptor of BC hearing. If usage of one of the components, or a combination of components, shows a change in FRF that correlates to the change in thresholds, then this could point to that being a good BC hearing indicator. When studying these differences between the triangle and frontal position in figure 4.8, it is observed that using the combination from Dobrev et al. would make the frequency response changes more closely correlate to the threshold changes compared to only using the y-component. By also studying the FRF for *HeadSim 2* for the positions, it is clear that these two positions have relatively big differences between the vibration components. Thus, this suggests that a combination of the vibration components is a better descriptor when changing the position this drastically.

This finding can also suggest the need for a combination of components when building a well-performing model for indicating BC hearing in *HeadSim 2*. The change in position between the triangle and frontal positions is very drastic, but the insight that a combination of spatial components can more accurately capture the threshold change could be applied even to small deviations. Thus, any proposed real-time implementable model that should be a good indicator for BC hearing regardless of stimulation position should most reasonably use all three spatial components.

5.2 Model Performance

The model created provided FIR filters. The filters' performance and connection to the appearance can be investigated as well as if any further improvements could be made with the existing data. As can be seen in figure 4.9, the filters are quite busy which can also be said about the filter output in figure 4.10. This appearance would be even more prevalent without any of the smoothing operations performed. Despite this, the overall trend of the output is matching the output, especially at the specific measured frequency points.

5.2.1 Data Use in System Identification

When analyzing the output from the combination model, both in figure 4.10 and 4.15, it is noted to have relatively big fluctuations with sharp notches and peaks. Since the proposed combination model is based on an LTI system approach, the output of a frequency is only dependent on the input with the same frequency. This could be an explanation for the appearance of the output since the three inputs are also fluctuating relatively much. In conjunction with matching the fluctuating inputs to a relatively consistent interpolated output, the determined filters could also have large fluctuations. This behavior can be noted in figure 4.9. By changing the position to one that is not present in the determination, the notches and peaks in the individual inputs could be slightly misaligned. This means, that if sharp notches or peaks would have moved slightly up or down in the spectrum for a specific component

compared to the positions used in determination, the output can be fluctuating relatively much. This is thus suggested to happen when the combination model is perfectly determined to the three positions used in the determination. In this way, the determined combination model can perfectly compensate for the notches and peaks in the components at the determination positions, but be sensitive to deviations when those changes when moving to another position.

In an attempt to minimize these fluctuations for the validation position, i.e. the triangle position, the filters were smoothed in third-octave bands. By performing this modification, it was observed that the estimated output was more closely following the original output at said validation position. Thus it was assumed that the model improved. However, this made the estimated output deviate from the original output corresponding to the positions used in the determination of the system, i.e. the mastoid, back, and frontal positions. This is seen because of that the determined frequency responses are modified, and after smoothing not able to exactly solve the equation at these positions. But overall, the performance at the triangle position improved with the third-octave moving mean, which led to the smoothing being implemented in the finalized filters. Further, by keeping the fluctuations in the transfer functions in the determined filters to a minimum, the needed FIR filter taps to describe the filter decreases.

Another aspect that could influence the fluctuations of the transfer functions for the determined filters is the use of data. Only the minimum amount of linearly independent equations was used when determining the system, i.e. only the measurements from three positions were used. This could make the determined solution, i.e. transfer functions, more specific to those positions and lack generalization. This could negatively impact the method when using new positions, methods to attach the actuator, or changing the head simulator. Thus, only using the minimum amount of data in the determination could make the model perform worse in validation. By making use of the data in a more efficient way or gathering more data, the model could be determined by finding the best solution to the problem, instead of finding the exact fit to this minimum amount of data.

These aspects can be seen as the issue of overfitting a model and getting a high variance, a concept more commonly associated with the field of system identification through machine learning [42]. Having a model overfitted to the system identification measurements will make the model lose performance at a validation position, i.e. having high variance. Regularization as a way to reduce the variance through minimizing overfitting is a concept that could also be applied when creating realizable systems, such as impulse responses in FIR filtering. When applying this point of view to the resulting combination model, measures such as using positions far apart, and using smoothing as a kind of regularization, can be assumed to improve the performance at the validation position since this minimizes variance. This can mean that by reducing the fluctuations in the transfer functions of the model the model is not able to perfectly fit the measured ”training” data, but is more generalized and performs better at the validation position.

Thus, since the combination model relies on acquired data, it is assumed that efficient use of this can improve performance. The existing data could be utilized in a

more efficient way by implementing a way of cross-validation. This would make sure that all of the measured positions contribute to a better determined, and hopefully more general, combination model. Moreover, increasing the amount of data could also increase the applicability and performance of a developed combination model. This could be done by either increasing the number of positions measured or by increasing the sample size of the head simulator and subject measurements. However, it is desired to keep the anatomy of the head simulator and the subject consistent. This could suggest that the same amount of head simulators as subjects used are needed.

5.2.2 The Importance of Phase

For the created model, it was somewhat assumed that the magnitude had a more significant role in the end result compared to the phase. The reason is that the phase shifts attributed to the vibration transmission through the skull were already present in the accelerometer data and accounted for, which was the input data for the model. Thus, the model is only responsible for the phase shift between the accelerometer, or cochlea, vibration to the perceived hearing or hearing threshold. Also, the model needs to make sure that the individual phase shifts for the components make the summation of the signal display the correct magnitude. For this, it was assumed that taking the average of all three space dimensions was a sufficient estimation since these phases were similar. In contrast, the method introduced by Dobrev et al. added the phases together in a much more sophisticated manner, which is part of the reason why this algorithm is difficult to implement in DSP hardware. Either way, both of these methods resulted in a similar listening experience in an offline environment. This might indicate that the more simple way to deal with the phase might be sufficient if it gives the opportunity to properly implement a combination model in DSP hardware. Nonetheless, the exact behavior and importance of the phase are something that can be investigated further.

5.3 Simulation in an Offline Environment

Analyzing the performance of the model could be quite challenging in some regards such as judging sound quality in an objective manner. Thus, comparing the power spectrum between simulation through the designed filters and the 10-s sound file was deemed to be a reasonable result to analyze. While listening to the sound file in figure 4.12a, and comparing it with the filtered sound in figure 4.12b, provided a good benchmark during development once the filters had been finalized it was hard to judge the performance in this manner. While the sound was certainly audible and perhaps similar to the original signal in some regards, it had more metallic characteristics. However, it is difficult to tell if the components had been combined into a realistic sound. On a similar note, while studying the power spectrum in figure 4.14 can provide information regarding what has happened with the signal, such as attenuation level at different frequencies, it is not exactly clear as to what a realistic result is based from this plot alone.

Thus, comparing against the method created by Dobrev et al. is interesting since this is found to be the only existing method that is adding the three spatial components together into one signal. While the goal is not necessarily to have the same end result as this method, it does provide some useful information in regards to analyzing the FIR combination method since judging the quality is as already mentioned quite difficult. The frequency response seen in figure 4.15 is interesting in this regard, as both methods behave quite similarly, which could also be heard when listening to the 10-s sound file processed via both of these methods. It was noticed that the sound quality got worse if the amplification was significantly different than the Dobrev et al. method, which meant it could provide some indication of performance by just comparing the output plots from both of the methods. However, having a lower amplification seen at higher frequencies did not affect the perceived sound quality as much, which is probably due to these frequencies are located outside of the speech spectrum. Furthermore, analyzing figure 4.16, it is interesting to see that despite the FIR model being based on the hearing threshold of the subject, the method from Dobrev et al. follows these thresholds perhaps even more compared to the combination model. However, compared to the y-component, which correlates less with the hearing thresholds, the proposed combination model could still provide a more realistic hearing perception since it correlates more with the hearing threshold changes.

While the two methods of combining the three axes seem to perform well in terms of following the hearing threshold, the overall audio quality is experienced as metallic sounding. This could be attributed to the higher amplification of some of the higher frequencies, such as a peak around 2000 Hz, as shown in the power spectrum in figure 4.14. This indicates that further processing once the three axes are combined into one signal might be necessary.

5.4 Real-Time Simulation Performance

When simulating in real-time, there are a lot more things to take into consideration compared to the offline setting. This mostly revolves around keeping latency and noise to a minimum in the chosen DSP hardware. The right and left side filters seen in figure 4.18 and figure 4.19 respectively are similar and the combined signal of both sides sounds similar, just as in the case of offline simulation. However, the gain of the two sides is slightly different since the FIR filter taps were adjusted to be less than one. Since the hearing threshold of the subject was in general the same on both ears, this difference in gain probably needs to be accounted for in order to have a realistic stereo sound. Thus, balancing the output from the two channels is something that needs to be performed.

In terms of latency, the complexity of the model is of great importance. FIR filters are very easy to implement in a DSP, and the complexity is governed by the number of filter taps. The power of the hardware dictates how long these filters can be. For the realized FIR filters, the amount of taps is well below what the hardware can handle, and the latency is not noticeable. On the other hand, using a different method of adding the components together could be much more demanding with

DSP hardware. For example, the particular hardware chosen did not accept any external algorithms created in another programming environment and thus had to be recreated in that particular development environment connected to the hardware. For FIR filters, this is not an issue, but for more complex algorithms this can be quite the endeavor. Also, the latency performance is not as easily analyzed before testing such an algorithm either. Thus, having a method based on FIR filters for real-time simulation seems to provide good results and is reliable in terms of latency.

When it comes to noise in a real-time setting, one thing to consider is to be in a quiet environment, since sound pollution from a noisy room will also be amplified by the DSP hardware. However, noise is still present, and there is a noticeable amount of noise when stimulated in real-time. Since noise will be present in all three axes, the combined signal could contain the noise from all three of them, which in extension could mean that noise is worse when listening to three-dimensional vibration data. Some higher frequency noise can be present, which can what is seen in figure 4.20 as higher amplification peaks in the higher frequency range. This can especially be seen at over 6000 Hz where the male voice frequency components are expected to be low.

The audio quality when listening in real-time is experienced as rather poor. The somewhat metallic quality present in the offline simulation is perhaps even more noticeable in the real-time setting. Studying the power spectrum in figure 4.20; there is plenty of amplification of higher frequencies that is probably not only due to noise, and was also present in the power spectrum of the noise-free offline simulation in figure 4.14. This could be one of the reasons why it sounds metallic since both of the simulated audio files consist of male voices. This means that the voices should not have components of significant power in this frequency range, which can be seen in the original sound files' power spectrums. There is, therefore, no obvious explanation for the origin of this behavior in the output from the combination model for these frequencies. Thus, the overtones from the voices could get amplified too much and thus resulting in a metallic-sounding output. This indicates that some further processing might be required after the combination model to increase the audio quality and thus make the sound more realistic.

6
Conclusion

Correctly and precisely simulating the BC hearing perception of a potential patient is a challenging task that is multifaceted. Insights are given by using threshold measurements from multiple locations at a subject's skull and correlating these to respective measurements at a head simulator. Changes in stimulation positions show a drastic change in orthogonal components not present in the measured hearing thresholds. Thus, only using one component could indicate being a less effective method. Comparing threshold differences to vibration frequency responses at the different positions indicates a better correlation for methods using a combination of the orthogonal component. This points to that all three orthogonal vibration components should be used when a model with the aim of correctly describing the BC hearing regardless of the position of stimulation is in demand.

The proposed combination model provides a way to efficiently combine the three orthogonal accelerometer readings into an audible signal. By basing the model on hearing thresholds and measurements of vibration propagation in the head simulator *HeadSim 2*, the model can be implemented through impulse responses and FIR filtering which enables real-time processing in DSP hardware. Comparing the proposed combination model with different methods, it can be seen that the proposed model follows the threshold differences between stimulation positions relatively well which suggests that this model provides satisfying results. This could point to a method closer to representing the BC hearing than any individual component could when spanning over multiple locations at the head simulator. The proposed combination model also shows promise since it is able to be implemented in hardware where the latency is regarded as unnoticeable. However, the complex nature of sound perception and interpretation makes the evaluation of the model challenging in an objective manner. Further investigating the performance could give insights into improvements in the model and its data.

Since the model is determined through system identification using measurements of inputs and outputs, the performance of the model is reliant on the amount of available high-quality data. Using a representative head simulator that changes the vibration components in accordance with the real subject, i.e. wave transmissions, is crucial to draw conclusions and determine the model. Uncertainties that might originate from audiometry, position measurement, and other disturbances can weaken the model. Using logarithmic smoothing and adjusting data points with assumed errors shows an improvement at the validation position. Further improvements might be found by increasing the amount of data used in the system identification. These efforts can both lessen the effect of disturbances or uncertainties and also suggest

a combination model which might give a generalized insight into how the vibration components contribute to BC hearing.

Using the combination model in offline simulation and real-time testing can show that the model efficiently combines the signals and outputs an audible output signal. However, the audio quality of this output signal might be influenced by aspects that make it deviate from a realistic case. The presence of noise and overamplification of overtones could give an output resulting in a metallic-sounding output. Thus, using the output from the combination model and further processing the signal could move the whole system closer to estimating realistic-sounding BC hearing.

The concept of developing the realizable combination model suggests a method that might move closer to BC hearing than currently implemented solutions in real-time processing. By having the combination model for each cochlea, it can efficiently produce a stereo sound in real time by processing the signals in parallel. By using every component of the sampled vibration data, a greater potential of *HeadSim 2* can possibly be reached. Thus, while the suggested model and method still could need refinement regarding sound quality, the proposed combination model could be an extra tool in order to bring the research and development closer to the experience of the end-user.

Bibliography

[1] World Health Organization. *Deafness and hearing loss*. (visited on 2023-05-11). 2023. URL: https://www.who.int/health-topics/hearing-loss#tab=tab_1.

[2] World Health Organization. *Deafness and hearing loss*. (visited on 2023-05-11). 2023. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[3] Cochlear Ltd. *About us*. (visited on 2023-05-16). 2023. URL: https://www.cochlear.com/us/en/about-us.

[4] Gerard J. Tortora and Gerard J. Derrickson. "Introduction to the Human Body". In: 11th ed. John Wiley & Sons, Inc, 2019. Chap. 12.6, pp. 294–301.

[5] Stefan Stenfelt. "Inner ear contribution to bone conduction hearing in the human". In: *Hearing Research* 329 (2015), pp. 41–51. ISSN: 0378-5955. DOI: https://doi.org/10.1016/j.heares.2014.12.003.

[6] American Speech-Language-Hearing Association. *Types of Hearing Loss*. (visited on 2023-05-11). 2023. URL: https://www.asha.org/public/hearing/types-of-hearing-loss/.

[7] Sabine Reinfeldt, Måns Eeg-Olofsson, Karl-Johan Fredén Jansson, Ann-Charlotte Persson, and Bo Håkansson. "Long-term follow-up and review of the Bone Conduction Implant". In: *Hearing Research* 421 (2022). ISSN: 0378-5955. DOI: https://doi.org/10.1016/j.heares.2022.108503.

[8] William Hodgetts, Dylan Scott, Patrick Maas, and Lindsey Westover. "Development of a Novel Bone Conduction Verification Tool Using a Surface Microphone: Validation With Percutaneous Bone Conduction Users". In: *Ear and Hearing* 39 (2018), pp. 1157–1164. DOI: 10.1097/AUD.0000000000000572.

[9] Bo Håkansson and Peder Carlsson. "Skull Simulator for Direct Bone Conduction Hearing Devices". In: *Scandinavian Audiology* 18 (1989), pp. 91–98. DOI: 10.3109/01050398909070728.

[10] Bo Håkansson, Fausto Woelflin, Anders Tjellström, and William Hodgetts. "The Mechanical Impedance of the Human Skull via Direct Bone Conduction Implants". In: *Med Devices (Auckl)* 13 (2020), pp. 293–313. DOI: https://doi.org/10.2147/MDER.S260732.

[11] Ashwin Kumaar Murali Dharan. *A Pilot Study on Developing an Artificial Head for Bone Conduction Devices Testing Purposes*. Mater Thesis, Chalmers University of Technology. 2021.

[12] Stefan Stenfelt and Richard L. Goode. "Transmission properties of bone conducted sound: Measurements in cadaver heads". In: *The Journal of the Acous-*

tical Society of America 118 (2005). ISSN: 0378-5955. DOI: https://doi.org/10.1121/1.2005847.

[13] Måns Eeg-Olofsson et al. "Transmission of bone conducted sound – Correlation between hearing perception and cochlear vibration". In: *Hearing Research* 306 (2013), pp. 11–20. ISSN: 0378-5955. DOI: https://doi.org/10.1016/j.heares.2013.08.015.

[14] Ivo Dobrev and Jae Hoon Sim. "Magnitude and phase of three-dimensional (3D) velocity vector: Application to measurement of cochlear promontory motion during bone conduction sound transmission". In: *Hearing Research* 364 (2018), pp. 96–103. ISSN: 0378-5955. DOI: https://doi.org/10.1016/j.heares.2018.03.022.

[15] Stefan Stenfelt. "Acoustic and physiologic aspects of bone conduction hearing". In: *Adv Otorhinolaryngol* 71 (2011). DOI: doi:10.1159/000323574.

[16] Roger B. A. Adamson, Manohar Bance, and James A. Brown. "A piezoelectric bone-conduction bending hearing actuator". In: *J Acoust Soc Am 1* 128 (2010). DOI: https://doi.org/10.1121/1.3478778.

[17] Bo Håkansson. "The balanced electromagnetic separation transducer: A new bone conduction transducer". In: *J Acoust Soc Am* 113 (2003). DOI: 10.1121/1.1536633.

[18] Amelia Bowlin and Sara Akbari. *Can you replicate Human Skin Properties using Artificial Materials for Bone Conduction Testing?* Mater Thesis, Chalmers University of Technology. 2022.

[19] Dytran Instrument. *MODEL 3293A*. (visited on 2023-05-22). 2017. URL: https://www.dytran.com/Model-3293A-Triaxial-Accelerometer-P2434/.

[20] William E. Hodgetts, Susan D. Scollie, and Ryan Swain. "Effects of applied contact force and volume control setting on output force levels of the BAHA Softband". In: *International Journal of Audiology* 45 (2006). DOI: https://doi.org/10.1080/14992020600582133.

[21] Nadia Verstraeten, Andrzej J. Zarowski, Thomas Somers, Daphna Riff, and Erwin F. Offeciers. "Comparison of the Audiologic Results Obtained With the Bone-Anchored Hearing Aid Attached to the Headband, the Testband, and to the "Snap" Abutment". In: *Otology & Neurotology* 30 (2009). DOI: 10.1097/MAO.0b013e31818be97a.

[22] Philip J Clamp and Robert JS Briggs. "The Cochlear Baha 4 Attract System – design concepts, surgical technique and early clinical results". In: *Expert Review of Medical Devices* 12 (2015). DOI: https://doi.org/10.1586/17434440.2015.990375.

[23] Francis F. Li and Trevor J. Cox. "Digital Signal Processing in Audio and Acoustical Engineering". In: CRC Press, 2019. Chap. 1, pp. 1–20. DOI: https://doi.org/10.1201/9781315117881.

[24] Horst Bessai. *MIMO Signals and Systems. [electronic resource].* Information Technology: Transmission, Processing and Storage. Springer US, 2005. ISBN: 9780387274577. URL: https://search.ebscohost.com/login.aspx?direct=true&db=cat07472a&AN=clec.SPRINGERLINK9780387274577&site=

eds-live&scope=site&authtype=guest&custid=s3911979&groupid=main&profile=eds.

[25] Anne Z. Saunders, Andrea Vallen Stein, and Nancy Lite Shuster. "Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition." In: Butterworth Publishers, 1990. Chap. 133. URL: https://www.ncbi.nlm.nih.gov/books/NBK239/.

[26] Svenska Audiologiska Metodboksgruppen (SAME-gruppen). "Handbok i hörselmätning". In: LIC förlag., 1990. Chap. 6, pp. 35–43.

[27] Svenska Audiologiska Metodboksgruppen (SAME-gruppen). "Handbok i hörselmätning". In: LIC förlag., 1990. Chap. 7, pp. 43–86.

[28] Ross J. Roeser, Michael Valente, and Holly Hosford-Dunn. "Audiology Diagnosis". In: Thieme Medical Publishers, Inc., 2000. Chap. 12, pp. 253–281.

[29] Stefan Stenfelt and Sabine Reinfeldt. "A model of the occlusion effect with bone-conducted stimulation". In: *International Journal of Audiology* 26 (2007). DOI: https://doi.org/10.1080/14992020701545880.

[30] Nur Annisa Rohaya Md Zakaria, Tengku Zulaila Hasma Tengku Zam Zam, and Raihana Rosli. "Hearing Threshold in Audiometry Testing: Pure Tone Versus Warble Tone." In: *International Journal of Allied Health Sciences (IJAHS)* 5 (2021). ISSN: 2530-2534. URL: https://journals.iium.edu.my/ijahs/index.php/IJAHS/article/view/459/644.

[31] Teresa A. Hamill and William H. Haas. "The relationship of pulsed, continuous, and warble extended-high frequency thresholds". In: *Journal of Communication Disorders* 19 (1986). DOI: https://doi.org/10.1016/0021-9924(86)90012-2.

[32] Jennifer J. Lentz, Matthew A. Walker, Ciara E. Short, and Kimberly G. Skinner. "Audiometric Testing With Pulsed, Steady, and Warble Tones in Listeners With Tinnitus and Hearing Loss". In: *American Speech-Language-Hearing Association* 26 (2017). DOI: 10.1044/2017_AJA-17-0009.

[33] Ivo Dobrev et al. "Influence of stimulation position on the sensitivity for bone conduction hearing aids without skin penetration". In: *International Journal of Audiology* 55 (2016). DOI: https://doi.org/10.3109/14992027.2016.1172120.

[34] Lizhe Tan. "Digital Signal Processing : Fundamentals and Applications". In: Elsevier Science & Technology, 2007. Chap. 1, pp. 1–3.

[35] Francis F. Li and Trevor J. Cox. "Digital Signal Processing in Audio and Acoustical Engineering". In: CRC Press, 2019. Chap. 4, pp. 55–80. DOI: https://doi.org/10.1201/9781315117881.

[36] Andrew P. McPherson, Robert H. Jack, and Giulio Moro. "Action-Sound Latency: Are Our Tools Fast Enough?" In: Conference: New Interfaces for Musical Expression (NIME). 2016. URL: https://www.researchgate.net/publication/315379106_Action-Sound_Latency_Are_Our_Tools_Fast_Enough.

[37] 3M. *3M™E-A-Rsoft™FX Earplugs*. (visited on 2023-05-22). 2013. URL: https://www.3m.com/3M/en_US/p/d/b00017634/.

[38] *SigmaDSP 28-/56-Bit Audio Processor with Two ADCs and Four DACs*. ADAU1701. Rev. 3. Analog Devices. Nov. 2005. URL: https://www.analog.com/media/en/technical-documentation/data-sheets/ADAU1701.pdf.

[39] Analog Devices. *EVAL-ADAU1467Z*. (visited on 2023-05-09). 2009. URL: https://www.analog.com/en/design-center/evaluation-hardware-and-software/evaluation-boards-kits/eval-adau146x.html#eb-overview.

[40] *Four ADCs/Eight DACs with PLL, 192 kHz, 24-Bit Codec*. AD1937. Rev. B. Analog Devices. June 2010. URL: https://www.analog.com/media/en/technical-documentation/data-sheets/AD1937.pdf.

[41] *SigmaDSP Digital Audio Processor*. ADAU1463/ADAU1467. Rev. A. Analog Devices. Nov. 2017. URL: https://www.analog.com/media/en/technical-documentation/data-sheets/ADAU1463-1467.pdf.

[42] Gianluigi Pillonetto, Tianshi Chen, Alessandro Chiuso, Giuseppe Nicolao, and Lennart Ljung. *Regularized System Identification: Learning Dynamic Models from Data*. Jan. 2022. ISBN: 978-3-030-95859-6. DOI: 10.1007/978-3-030-95860-2.

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY