The benefits and user experience of hearing aids controlled by eye gaze of the user

Lucca Julie Nellemann

Master Thesis
in cooperation with Oticon A/S and Eriksholm Research Centre

Fall Semester 2019

Aalborg University
Engineering Psychology

This report is compiled in LaTeX. Additionally MATLAB is used to run the experiment and Rstudio, SPSS Statistics, and Matlab is used to analyze present the results.

Department of Electronic Systems
Engineering Psychology
Fredrik Bajers vej 7
9220 Aalborg

AALBORG UNIVERSITY
STUDENT REPORT

Title: The benefits and user experience of hearing aids controlled by eye gaze of the user
Theme: Master thesis

Author: Lucca Julie Nellemann
University supervisor: Lars Bo Larsen
Company supervisors: Poul Hoang & Martha Shiell

Number of pages: 97

Project period: Fall Semester 2019
Date of completion: January 7th 2020

Abstract

A problem with hearing aids is that they do not always reduce the noise correctly, which can make it hard to follow a conversation in noisy surroundings. This master thesis seeks to investigate if eye gaze steering of a hearing aid can benefit the hearing aid user by testing two forms of eye gaze steering on a concept level; hard and soft eye gaze steering. An experiment is conducted with 13 hearing aid users answering questions by following a conversation during four different conditions; a familiarization round and a no, hard, and soft eye gaze steering condition. For every condition, participants answered a NASA Task Load Index (TLX) questionnaire for measuring workload, and after the experiment, an exit-interview was conducted. When calculating the percentage of correct answers during each condition the results show the familiarization round to have the most correct answers, followed by the hard eye gaze steering, soft eye gaze steering, and no steering with least correct answers. A significant difference is found between all conditions, except the hard and soft eye gaze steering. When measuring workload the only difference is found between the familiarization round and the three other conditions, where the familiarization is scoring lowest. The eye gaze steering of a hearing aid seems to work by reducing noise that participants do not want to listen to, and when asked about their preferences, participants preferred the hard eye gaze steering.

Resume

Et problem der er med høreapparater i dag er, at de ikke altid reducerer støjen korrekt, hvilket kan gøre det svært at følge en samtale i støjende omgivelser. Dette kandidatspeciale stræber efter at undersøge om øjestyring af høreapparater kan have en fordelagtig indvirkning på brugen af høreapparater. Dette gøres ved at teste to typer af øjestyringsalgoritmer på konceptplan; en hård og en blød øjestyringsalgoritme.

Et eksperiment er udført med 13 høreapparats brugere, der ved at følge en samtale på en TV-skærm skal svare på spørgsmål under fire forskellige konditioner; en familiariseringsrunde og en kondition med henholdsvis ingen, hård og blød øjestyring. For hver kondition svarer testpersonerne efterfølgende på et NASA Task Load Index (TLX) spørgeskema, der måler arbejdsbyrden. Efter eksperimentet er der afholdt et exit-interview.

Når procentfordelingen af antal rigtige svar udregnes og analyseres, viser resultaterne, at der bliver svaret flest spørgsmål rigtigt under familiariseringsrunden. Derefter følger hård øjestyring, blød øjestyring og ingen øjestyring med færrest korrekte svar. En signifikant forskel er fundet mellem alle konditionerne, undtagen hård og blød øjestyring. Når arbejdsbyrden bliver analyseret er den eneste forskel fundet mellem familiariseringsrunden og de tre andre konditioner, hvor familiariseringen scorer lavest. Øjestyring af et høreapparat ser ud til at virke, da uønsket støj bliver reduceret i den retning testpersonerne ikke kigger. Når testpersonerne bliver spurgt om hvilken form for øjestyring de foretrækker, er det klart hård øjestyring der er mest foretrukket.

Preface

This report is composed by Lucca Julie Nellemann during the 4th semester of the master's programme in Engineering Psychology at Aalborg University. The 4th semester is dedicated to the Master Thesis, which in this case is a study about how to improve hearing aids in cooperation with Oticon and Eriksholm Research Centre, which is a part of Oticon. This report strives to document the project during the literature search, user experiments, data analysis, and reflection upon the results.

This master thesis project is part of a bigger project on Oticon and Eriksholm Research Centre, which is why both company supervisors have been a big part of the project. The program used in the experiment is written by one of the company supervisors, Poul Hoang, with assistance from Martha Shiell and Sergi Rotger Griful.

For citation the report employs the Harvard method. If a citation is not present in tables or figures, they are produced by the author.

Aalborg University, January 7th 2020

Lucca Julie Nellemann
<ljne14@student.aau.dk>

Contents

Chapter 1

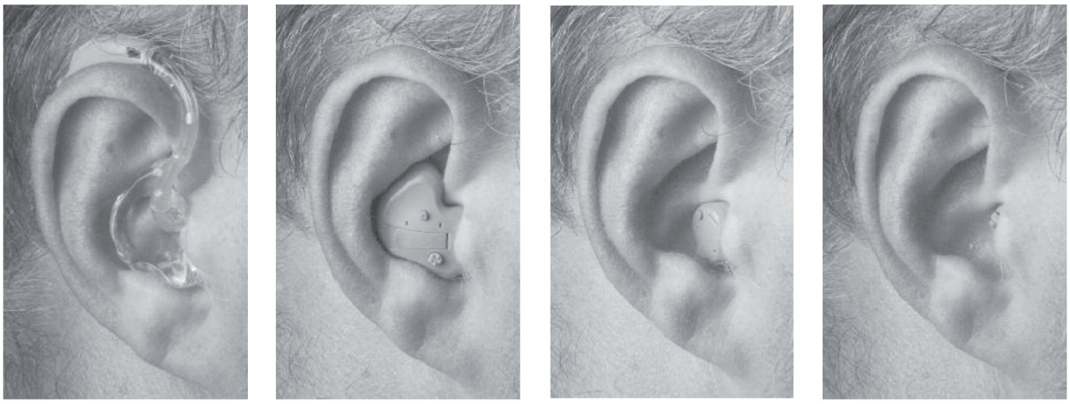
Introduction

The human hearing is one of the most important senses since it enables us to communicate through spoken language and in general makes us aware of our environment. Unfortunately, a lot of people are either born with a hearing impairment or develop it over time, which influences their life greatly. WHO estimates that around 466 million people worldwide have disabling hearing loss, which is around 5% of the world's population, [World Health Organization, 2019]. Furthermore, it is estimated that 900 million people will have disabling hearing loss around the year 2050.

Hearing impairment is unpleasant and can affect the everyday life of both the hearing impaired and the people next of kin, [Elberling and Worsøe, 2007, pp. 37-42]. It is usually everyday situations that become more difficult, such as understanding the family members and following conversations. Furthermore, the hearing impaired have difficulties adjusting the loudness of his or her own voice, because they cannot hear it properly anymore, [Elberling and Worsøe, 2007, p. 37]. This often results in the hearing impaired speaking very loudly, which most likely will bother people at the other end of the conversation. An even worse situation is it to have a hearing impairment in a noisy environment, [Elberling and Worsøe, 2007, pp. 40-41]. A classic cocktail party or another form of social gathering of people is especially difficult because the hearing impaired has difficulties locating the desired speaker, and separating speech from noise, which usually results in a poor experience being at these social events. A hearing impairment will therefore greatly affect everyday life, and especially during social interactions.

One solution to a hearing impairment could be to use a hearing aid. Even though hearing aids cannot restore a normal hearing, they can greatly improve the perception of surrounding sounds and communication with other people, [Elberling and Worsøe, 2007, p. 72]. In the world of hearing aids, various types and designs exist, [Elberling and Worsøe, 2007, pp. 57-59]. There is the Behind-The-Ear (BTE) hearing aid, which is placed behind the pinna and leads the sound to the ear canal through a small plastic tube, see figure 1.1a. Furthermore, there are hearing aids, which are placed inside the ear canal: In-The-Ear (ITE), In-The-Canal (ITC), and

Completely-In-Canal (CIC). The ITE is placed in the ear canal, but also fills out a part of the outer ear, see figure 1.1b. ITC and CIC are both placed inside the ear canal, see figure 1.1c and 1.1d, where the CIC is almost invisible. Which hearing aid a user needs depends on the extent of their impairment, and which activities are meaningful for them in their everyday life, [Elberling and Worsøe, 2007, pp. 72-74].



(a) BTE hearing aid. **(b)** ITE hearing aid. **(c)** ITC hearing aid. **(d)** CIC hearing aid.

Figure 1.1: The four different kinds of hearing aids, [Elberling and Worsøe, 2007, pp. 56 and 58].

Even though a lot of different hearing aids are available, not all people with hearing impairment find them pleasurable or comfortable to use, [Hart et al., 2009]. This is often due to the undesired amplification of background noise that hearing aids do not always identify as noise. When both the target signal and background noise are enhanced, the hearing impaired can experience problems distinguishing between the sounds, [Elberling and Worsøe, 2007, pp. 47-51]. Furthermore, when a hearing impaired is not used to wear a hearing aid, they can be overwhelmed by all the sounds that are suddenly present, caused by the amplification, [Elberling and Worsøe, 2007, p. 55]. It takes time to get used to a hearing aid, usually several months, during which it is also important to get the hearing aid adjusted for the best fit, [Elberling and Worsøe, 2007; Betkowski, 2018]. Lastly, users expect a hearing aid to restore the hearing completely, which is not possible.

Even though it is not possible to restore the hearing completely using hearing aids, it is possible to improve the noise reduction algorithm, so hearing aids will be better to separate noise and desired sound to increase speech intelligibility. When the noise has been distinguished from the desired sound, noise reduction can be made, followed by the specific hearing loss compensation that the user needs, [Launer et al., 2016, p. 97]. If the noise reduction technology improves, it is likely that the user experience will also improve, which in the end can motivate more users to keep using their hearing aid and more hearing impaired to begin using hearing aids.

Chapter 2

Literature review

The following chapter will explain the theory and previous research related to this Master thesis, including how the human hearing function, which kind of hearing impairments is known, how hearing aids work, and how those can be improved.

2.1 The human hearing

The human ear consists of the outer ear, the middle ear, and the inner ear, which can be seen on figure 2.1, [Moore, 2013, pp. 23-24]. The outer ear consists of the pinna, which is the visible part, and the auditory canal. The pinna is important because it modifies incoming sound and helps localize sound. Sound is vibrations in the air, and when it propagates through the ear canal, it eventually reaches the eardrum, which then starts to vibrate. The middle ear consists of three small bones, the malleus, incus, and stapes, which transmit the vibrations from the sound to the inner ear through the oval window. The inner ear consists of a spiral-shaped cochlea, which overall translates mechanical vibrations into electrical signals, and the semi-circular canals, which are concerned with balance. The basilar membrane is placed inside the cochlea, and when sound enters the cochlea, the basilar membrane vibrates relative to which frequencies the sound has. The basilar membrane is connected with the hair cells of the ear, where the role of the outer hair cells are to actively influence the mechanics of the cochlea, hereunder produce high sensitivity and sharp tuning, [Moore, 2013, p. 35]. The inner hair cells act to transduce mechanical movement into neural activity, so the brain understands the sound coming into the ear. When a person is said to have a hearing impairment, it is usually somewhere in the outer, middle, or inner ear the problem occurs.

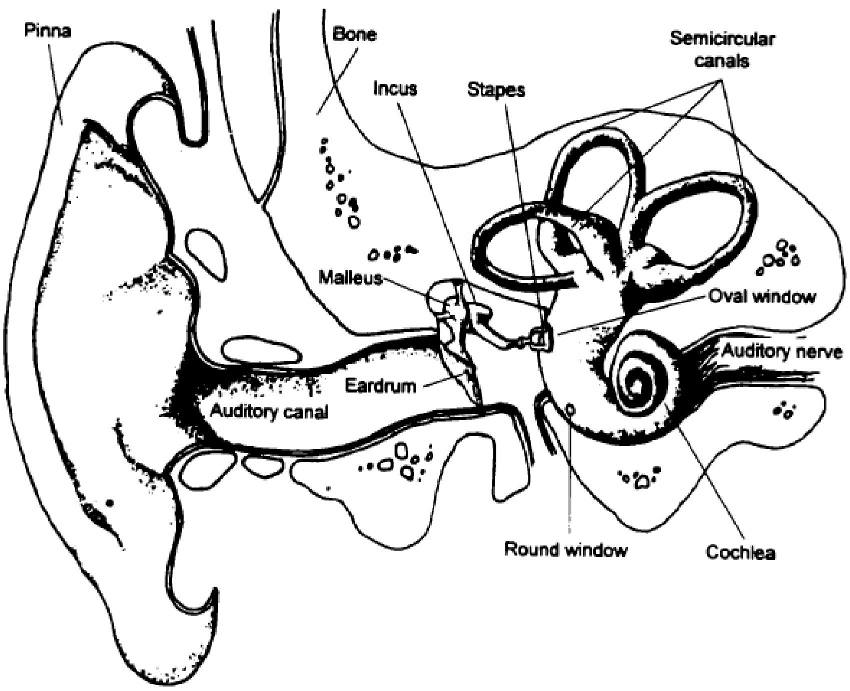


Figure 2.1: Illustration of the auditory system showing the outer ear, the middle ear, and the inner ear, [Moore, 2013, p. 23]

2.1.1 How people distinguish between sounds

In a social situation with several people talking at the same time, for instance at a social gathering at a restaurant or a cocktail party, it can be hard to follow a conversation due to the increased noise. However, the human mind is usually able to follow one stream of information, while everything else is perceived as background noise, defined as the "cocktail party effect", [Cherry, 1953]. This ability often stems from the different input in the two ears, binaural cues, which makes it possible to locate sound, [Brungart et al., 2001]. One binaural cue is the interaural time differences (ITD), which is defined as the difference in time between a sound wave arriving at one and the other ear, [Moore, 2013, pp. 247-250]. Another binaural cue is the interaural level differences (ILD), which is defined as the difference in sound pressure level entering the two ears. Since the head "shadows" for sound with short wavelengths, ILD is often used to localize high-frequency sounds, while ITD is used to localize low-frequency sounds. Despite the human ability to follow a stream of information at a social gathering, problems with hearing can occur due to problems with binaural cues or other psychophysical phenomena.

One phenomenon is that the stream of information, or sound from the target, can be masked when noise is present, [Moore, 2013, pp. 67-68]. Masking can be either energetic or informational, [Brungart et al., 2001]. Energetic masking is the more traditional form of masking, where the desired sound is masked by noise in such a way, that you cannot separate the two sounds from each other. This occurs when the background noise and desired sound are present at the same time or with very little time apart, [Moore, 2013, pp. 67-68]. Masking often occurs, when the noise and desired sound consists of frequencies close to each other, but can also occur when the masking sound is very loud and therefore contains enough energy to mask the desired sound. The concept of masking can be seen on figure 2.2 and 2.3, where figure 2.2 shows two tones with frequency far enough from each other to be distinguished as well as loudness that does not affect each other, and figure 2.3 shows how a high-frequency sound can be masked by a lower frequency sound. Another form of masking is called informational masking and refers to situations, where the masker and target sound are both audible, but the listener is unable to separate the information because both inputs are very similar, [Brungart et al., 2001]. At a cocktail party, both energetic and informational masking can occur. Speech contains frequencies close to each other, while the loudness of other speech signals can mask the speech signal you are interested in, which can result in energetic masking. Furthermore, speech signals at a cocktail party can also be very similar with respect to the context, which can create information masking of the desired speech signal.

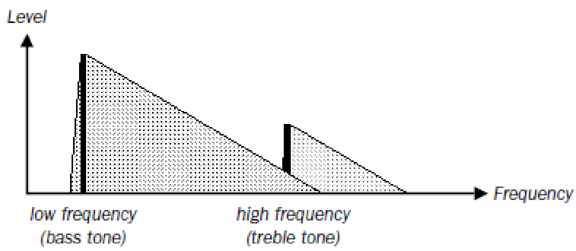


Figure 2.2: A high and low frequency sound that can be distinguished from each other, [Elberling and Worsøe, 2007, p. 48].

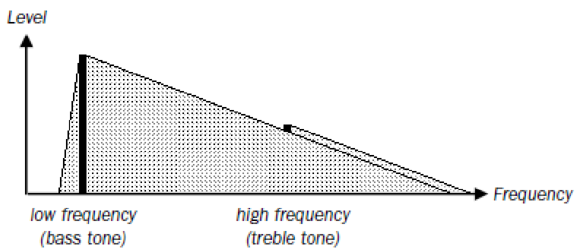


Figure 2.3: A low frequency sound that mask the higher frequency sound, [Elberling and Worsøe, 2007, p. 48].

2.1.2 Hearing impairments

Hearing impairments can stem from a range of problems in the ears, but they can be broadly categorized into two types, [Moore, 2013, p. 62]. The first type is conductive hearing loss, which usually occurs when there is a problem in the middle ear, but sometimes also in the outer ear. This can, for instance, be viscous fluid that builds up in the middle ear as a result of an infection or too much wax in the ear canal. A conductive hearing loss reduces the transmission of sound to the cochlea, resulting in loss of information about the volume of sounds.

The second type of hearing loss is called sensorineural hearing loss and most commonly arises from damage in the cochlea, [Moore, 2013, p. 62]. This weakens the signals to the brain and results in loss of information about volume, clarity, and the ability to distinguish between sounds is weakened.

The five dimensions of hearing

Even though problems with the hearing usually stem from one of the two mentioned hearing impairment types, it can also be a mixture of the two types. [Elberling and Worsøe, 2007, pp. 43-55] describes five dimensions of the hearing, which can all have an important role when looking at a hearing impairment. These five dimensions are described as follows:

- Sensitivity - The sensitivity of the ear is determined by the hearing threshold, [Elberling and Worsøe, 2007, pp. 43-44]. This describes at which sound pressure level you are just able to detect a sound. A hearing-impaired with problems in this dimension will have a higher hearing threshold than normal hearing people.
- Dynamic range - The dynamic range is the range between the hearing threshold and the loudest sound that can be tolerated for a short moment, [Elberling and Worsøe, 2007, p. 45]. This dynamic range can be either large or small, depending on the hearing threshold. Usually, the uncomfortable loudness level changes only slightly through a lifetime, which means that when the hearing threshold is raised, a smaller dynamic range is available.
- Frequency resolution - Frequency resolution is closely connected with frequency selectivity and concerns frequencies that mask each other on the basilar membrane, [Elberling and Worsøe, 2007, pp. 47-49]. There is a tendency for lower frequencies to mask higher frequencies, which can make it hard to distinguish sounds with frequencies close to each other. For a hearing impaired, the vibration pattern on the basilar membrane can be wider than normal, which makes it even harder to distinguish sounds, even when they are not very close to each other.
- Temporal resolution - Temporal resolution is closely connected with masking and is concerned with sounds that are closely following each other, [Elberling and Worsøe, 2007, pp. 49-51]. When the first sound has sent information to

the central nervous system it can take some time before the system is able to handle another sound or information. Therefore the first sound can disturb the following sound, which results in the listener not hearing or understanding the second sound. For a person with hearing impairment, the temporal resolution can be reduced.

- Hearing with two ears - This dimension concerns binaural cues, [Elberling and Worsøe, 2007, pp. 51-52]. Usually, it is possible to localize sounds in the surroundings due to the two ears, which can make it easier to understand or follow one stream of information. When the hearing is impaired, the brain does not receive the correct binaural information, which will reduce the ability to localize sound and in the worst case reduce speech intelligibility when there's disturbing background noise.

Depending on where in the five dimensions the problems occur, it can be either a conductive hearing loss or a sensorineural hearing loss. If problems occur on several dimensions, then the hearing impaired will probably have a mixture of the two types of hearing impairment.

2.2 Hearing aids

For the most common types of hearing loss, hearing aids can greatly improve the hearing, including communication with other people and the experience of surrounding sounds, [Elberling and Worsøe, 2007; Launer et al., 2016]. Depending on where in the ear, or in what dimension, the problem occurs, different compensations can be made, [Elberling and Worsøe, 2007, pp. 72-74]. If the hearing impaired has problems in several dimensions of the hearing, compensating for one problem in one dimension may affect another dimension. The settings in the hearing aid are therefore a compromise between the user's wishes on how the hearing aid should compensate. Furthermore, a fundamental problem is that even though hearing aids can compensate for a lot of impaired functions in the ear, the sound still has to pass through the damaged part of the ear, which brings some limitations. To give the hearing impaired the best chance of understanding speech, the hearing aid can help by reducing the noise in the environment.

Noise reduction in hearing aids is done before compensating for the hearing loss, [Launer et al., 2016, p. 97]. When sound enters the hearing aid, the direction of the target sound is determined. Sounds coming from behind or the side of the user are usually classified as noise, meaning that the hearing aid will compensate in the nose direction of the user. The signal is cleaned by reducing the noise coming from the hearing aid itself, as well as noise coming from behind or the side of the user, wind noise, or similar. All of this is done to get a clearer target signal when reaching the hearing compensation system.

Directional microphones and beamformers in hearing aids

The direction of the target sound is often determined by the nose direction of the user and sound can be enhanced in the target direction by using directional microphones. When it is not possible to have a remote microphone at the sound source, the best way to improve speech intelligibility is by using directional microphones, [Launer et al., 2016; Dillon, 2012]. Directional microphones can be constructed from a single microphone with more than one entry, where the time difference in the entry signals can help determine where the sound is coming from, [Dillon, 2012, p. 199]. Otherwise, directional microphones can be constructed by using more than one microphone and the output from these microphones. A subcategory to directional microphones consisting of multiple microphones is a beamformer. A beamformer is an algorithm which for instance can enhance a sound in a specific direction, using more than one microphone. It should be noted that a directional microphone consisting of a single microphone and several ports is not a beamformer.

Even though using directional microphones is one of the best ways to improve speech intelligibility, there are also disadvantages with hearing aids that carry these microphones. Most important is the problem of desired sound from behind or the sides of the user, which is reduced, [Launer et al., 2016; Dillon, 2012; Hart et al., 2009]. This can, for instance, happen when the hearing impaired is driving a car with passengers. Listening to the passengers sitting on the side or in the backseat can be hard since the directional microphones are picking up sound in the nose direction of the user. In these specific situations, omnidirectional microphones will be a better fit for the hearing aid user. However, hearing impaired will in general benefit from directional microphones, since they need a better signal-to-noise ratio (SNR) than normal hearing people, which directional microphones can give them, [Dillon, 2012, p. 198].

Directional microphones or beamformers can only be used in larger hearing aids, which is why they are mostly used in BTE or ITE hearing aids. In smaller hearing aids there is not enough room for the necessary spacing between microphones or ports, [Dillon, 2012, p. 201]. Because directional microphones are placed in either BTE or ITE hearing aids, there can be some limitations in determining the right direction of the sound and furthermore enhancing sound in the correct direction, [Launer et al., 2016, pp. 111-112]. This is due to the pinna shadowing for important information about the directivity when the hearing aid is placed either behind the ear or inside the ear.

Besides the already mentioned limitations of hearing aids with directional microphones, or directional hearing aids, there can be additional problems when using head direction, or nose direction, as steering of a beamformer in a multi-talker scenario. Hart et al. [2009] mentions that the hearing aid steered with head direction is not measuring where the user is actually looking. A hearing aid user can therefore turn the head towards person A, but actually be listening to and looking at person B. In a multi-talker scenario, it is also possible that the hearing impaired is turning their head to be in between person A and B while the eyes are jumping back and forth between person A and B. In this case, none of the correct signals will be enhanced and in the worst case be determined as noise because the hearing aid can have problems with knowing the desired target sound when. The same tendency is seen by Harrison [2018], who shows that hearing aid users do not turn their head all the way to a speaker on the side of them, even though they may be looking at the speaker.

2.2.1 How to improve hearing aids

As described, problems with directional microphones can occur due to the placement of the hearing aid and the pinna shadowing for incoming sound. This results in problems with determining the direction of sound and problems with enhancing sound or reducing noise in the correct direction. Furthermore, problems with steering the beamformers with head or nose direction, also called head steering, can occur, because the hearing impaired is not always turning their head all the way to the person they are listening to, even though they are looking in that direction. To improve hearing aids as they are today, it could be beneficial to look at how eye gaze may be used to determine, or in combination with directional microphones help determine, in which directions the noise should be reduced or desired sound enhanced.

Eye gaze in conversation

Vatikiotis-Bateson et al. [1998], Buchan et al. [2008], and Vatikiotis-Bateson et al. [1994] have investigated how people look at a talking person in noisy environments. It is concluded that people tend to look in the direction of the speaker and benefit from having an audiovisual perception of the speaker, especially in noisy environments. Furthermore, it is concluded that the duration of eye gazes at the target is longer when there is noise present, [Vatikiotis-Bateson et al., 1998, 1994]. Research from Vertegaal et al. [2001] showed that people look at the speaker 88% of the time when they are listening, and look 77% at the person they are directing their speech to. This research is conducted with normal hearing participants, but people with hearing impairment might look even more at the speakers when they are trying to listen, especially because it will give them the benefit of also lip-reading, [Elberling and Worsøe, 2007, p. 41]. Since eye-gaze is described as an excellent predictor of conversational attention, [Vertegaal et al., 2001], and eye contact furthermore is a natural human response in a social environment, it seems fair to assume that eye gaze tracking to support or steer beamformers to some extent can have a beneficial contribution in noise reduction technology for hearing aids.

Eye gaze steering of beamformers

Previous research have investigated eye gaze steering of beamformers and how it could benefit hearing aid users, among these Kidd et al. [2013], Hart et al. [2009], Harrison [2018], and Best et al. [2017]. Kidd et al. [2013] used an acoustic beamforming microphone array coupled with an eye-glass mounted eye tracker to test the concept of steering a hearing aid with eye gaze. This preliminary test revealed that the concept worked, even though no wearable prototype was made. Hart et al. [2009] tested how amplification of sound worked when using eye gaze to steer compared to either pointing at the sound source, pressing a button linked to the sound source, or a control, that presented an omnidirectional hearing aid. Results showed that eye gaze steering was 73% faster than pointing at the sound source and 58% faster than button selection. In terms of recalling the presented material, participants did 37% better using the eye gaze steering condition than pointing condition. Furthermore, eye gaze steering proved to be 54% better than button selection and 80% better than the control condition. Eye input was actually significantly better than all other three options when it came to participants recalling the material. Hart et al. [2009] also tested the user experience and found that participants rated eye gaze steering highest in categories as "easiest", "most natural", and "best overall".

Harrison [2018] investigated eye gaze steering and compared it with head steering together with Eriksholm Research Centre. Both steering forms were tested in a multi-talker scenario. Results showed that both head steering and eye gaze steering improved hearing performance of the hearing impaired more than their ordinary hearing aids. However, eye gaze steering was found to be faster and more precise than head steering, especially when it came to directing the attention towards speakers on the side of the test participant.

Best et al. [2017] investigated how eye gaze steering of a beamformer worked in both fixed and dynamic settings, with respect to one day steering a hearing aid like this. They concluded, that a visually guided hearing aid can provide benefits for users listening to a fixed target, but also concluded that the benefits were reduced under dynamic conditions. This is due to the user needing to shift the eye gaze, which can take around 500 ms from initiation to completion of eye movement. Therefore first words of sentences can be missed if the user has to first shift the eye gaze back and forth.

The eye gaze steering described in previous research works theoretically by enhancing the sound in the direction of the eye gaze to any given time. This will be referred to as hard eye gaze steering of beamformers.

Even though previous research confirms that eye gaze steering of beamformers can be a good way to improve hearing aids, this hard eye gaze steering also has some limitations. Many beamforming methods require the direction of arrival of the desired sound source to be known in order to achieve optimal noise reduction performance, [Hoang et al., 2019]. If the user does not look at the person talking all the time, it will result in dropouts of the desired speech, because the beamformer will enhance in the direction of the eyes and then classify the actual desired sound as noise. Furthermore, Best et al. [2017] concluded that in a multi-talker scenario, it can be hard to shift eye gaze quickly enough to catch the first words of a new talker. Chakrabarty et al. [2015] also argues that the performance of hard beamformers degrade in noisy environments because these might lead to distortion of sound and degradation of filter performance, which can make the desired speech harder to understand.

Instead of having a hard beamformer enhancing sound in the direction of the user's momentarily eye gaze, it has been proposed to use a Bayesian beamformer, which should be more robust to the errors there are when determining the direction of sound, [Hoang et al., 2019; Chakrabarty et al., 2015]. This Bayesian beamformer measures eye gaze over a time period and determines how much enhancement the signals in the different directions should have, according to how long the eye gaze was pointing in that direction, Chakrabarty et al. [2015]. When the Bayesian beamformer is tested using microphone signals, it outperforms other state-of-the-art beamformers, both in low and high SNRs, when the target direction is uncertain, [Hoang et al., 2019]. With that in mind, it could be beneficial to investigate how a Bayesian beamformer of eye gaze steering could work to improve the experience with hearing aids. Since a Bayesian beamformer collects eye gaze data over time, it might not be dependent on the user shifting the gaze immediately for one talker to the next in a multi-talker conversation. Furthermore, it leaves room for the hearing impaired to look away from the speaker for a brief moment and will most likely reduce the distortion of sound when the gaze is shifted in a noisy environment. The Bayesian approach will be referred to as soft eye gaze steering of beamformers.

Chapter 3

Project focus

As described in section 2.2.1, the concept of hard eye gaze steering of beamformers to improve noise reduction in hearing aids has been tested with promising results. However, there seem to be several issues with hard eye gaze steering of beamformers, which might improve with a Bayesian approach to eye gaze steering of beamformers. This master thesis will in corporation with Oticon and Eriksholm Research Centre aim to investigate the benefits of hard and soft eye gaze steering of beamformers. Furthermore, it will aim to investigate how these concepts work from a user perspective. This leads to an overall research question:

"How does eye gaze steering of beamformers help the hearing aid user, and is hard eye gaze steering or soft eye gaze steering the better choice?"

To understand how the two forms of eye gaze steering help the hearing aid user, they should be tested against each other in a noisy situation. Furthermore, it is interesting to investigate how both hard and soft eye gaze steering works compared to no steering in the same noisy situation. As described by Vatikiotis-Bateson et al. [1998] and Vatikiotis-Bateson et al. [1994], eye gazes are longer in noisy environments, which might mean that soft eye gaze steering of a beamformer will somewhat act like hard eye gaze steering if the gaze is kept at the talker for longer periods under noisy environments. However, it is assumed that soft eye gaze steering works better than hard eye gaze steering, and is furthermore preferred by the user in multi-talker scenarios because the user should be able to move the eye gaze more freely without loosing the desired sound in the directions they are looking most of the time.

When the concept of hard and soft beamformer should be implemented in hearing aids, it will work by reducing the noise in the direction the user is not directing his or her eye gaze. In this master thesis, working noise reduction algorithms will not be implemented in hearing aids. Instead, the concepts will be tested by enhancing and reducing sound in the direction that the participant is looking and not looking, respectively, according to which form of eye gaze steering is used.

For this project, several hypotheses are set up to be tested.

3.1 Hypotheses

To investigate the benefits of hard and soft eye gaze steering of beamformers, the following hypotheses are set up:

- **Hypothesis 1:** Eye gaze steering is better than no steering in a noisy environment when measuring
  (a) Speech intelligibility
  (b) Workload
- **Hypothesis 2:** Soft eye gaze steering is better than hard eye gaze steering in a noisy environment when measuring
  (a) Speech intelligibility
  (b) Workload
- **Hypothesis 3:** When asked about their opinion, users will prefer soft eye gaze steering of a beamformer over hard eye gaze steering of a beamformer and no steering.

To investigate the research question and the hypotheses, an experiment is set up.

Chapter 4

Experimental setup

To investigate the benefits of hard and soft eye gaze steering of beamformers, an experiment is conducted. This experiment will gather both quantitative and qualitative data about the participants and their behavior to answer both the research question and the different hypotheses. Throughout the experiment, participants with hearing aids will therefore be exposed to the following:

- Hearing in noise test (HINT) to gather information about their hearing threshold in noise.
- Introduction to the system and the materials used.
- Four conditions; familiarization round, no steering, hard eye gaze steering, and soft eye gaze steering.
- Tasks for every condition, including speech intelligibility questions and a NASA Task Load Index (TLX) questionnaire to measure workload.
- An exit-interview.

Throughout this chapter, every aspect of the experiment will be described. This will include the mentioned points, but also the material used, the software, the physical setup, the participants, and the procedure of the experiment.

People with hearing aids often have very different hearing impairments, and since their individual differences might be enough to give different results on the different forms of eye gaze steering conditions, a within-subject design is chosen.

4.1 Audiovisual material

To investigate how people with hearing impairments use their eye gaze in a multi-talker scenario and how eye gaze steering can help, it is important to present a multi-talker scenario for the participants. It was discussed to have real people be present in the room during the experiment, and enhance the sound from them in the direction the participants were looking because this would make the scenario very natural. Due to an uncontrollable variable such as the present talkers always had to speak with the same loudness, this idea was discarded. Instead, it was decided to show videos of people talking with each other on a big curved TV screen, which could show the talkers in live sizes.

Eriksholm Research Centre has previously recorded twelve three-talker scenario videos with four (two male, two female) danish actors. An example of these videos can be seen on figure 4.1 and in appendix A.1. On the videos, one of the actors is having a monologue on either the right or left side of the screen, approximately 20 degrees from the middle talker on both sides. He or she is telling a random story, while a dialogue is taking place between the two other actors. In the dialogue, they are both given a picture of a landscape. The pictures differ in twelve places, and the actors in the dialogue are told to talk about the pictures and locate the twelve errors. All actors were wearing hands-free microphones placed close to their mouths, so their clean speech signal was recorded synchronized with the video.

Figure 4.1: An example of the audiovisual material filmed by Eriksholm Research Centre. Here the person to the left is having a monologue, while the middle and right person is having a dialogue.

It was decided to use videos that 1) always had the monologue be the same gender as one of the talkers in the dialogue, 2) always had the talkers' mouths visible and not covered by the picture they were holding, and 3) to balance the talkers as much as possible. From the twelve videos recorded, four videos were chosen to be useful in this experiment, due to these selection criteria.

For the talkers to have the same amount of loudness, all speech signals were normalized. In one of the videos (video 9), the middle talker had his microphone too close to his mouth, which resulted in clipping of his voice. This audio-signal was run through different restoration algorithms as de-clipping and A-weighting for it to be useful. As all other speech signals, this speech signal was afterward normalized to an RMS value of 0.01.

4.1.1 Speech intelligibility questions

To answer hypothesis 1(a) and 2(a), speech intelligibility should be measured. When talking about speech intelligibility, it is usually the recognition of words and how many words are recognized. However, when using material with a natural conversation, as the audio-visual material is, it is hard to measure the correct amount of words. Instead, it was decided to develop questions that could be answered by seeing small clips from the videos. During this master thesis, speech intelligibility will therefore be referring to the percentage of right answers participants could give when asked different questions.

69 video clips lasting between 18 to 30 seconds and questions, which were answered between 12 to 25 seconds into the video clip, were developed from listening to the dialogue. These questions were then presented to a normal hearing person for evaluation. Some questions did not make the cut, because they were hard to understand or the answer was unclear. Furthermore, it was decided to always use different video clips and have the overlapping of video clips as minimal as possible. The total count of video clips and questions from the four videos ended up at 45, divided into 11 questions per condition. The questions can be seen in appendix B.

4.2 Signal to noise ratio

To understand the benefits of eye gaze steering in noisy environments, background noise has to be present. For this experiment pedestrian noise was used as background noise because it sounded like being at a noisy place with other people in the background. Babble noise was also considered but deselected because it seemed weird to have that many voices of talkers present when only three talkers were visible.

Since the no steering condition works as a control condition, it is beneficial that participants score around 50% in speech intelligibility, or correct answers, in this condition. This is because they then have room to improve with the eye gaze steering conditions, while they are not so bad as this condition that they give up. Furthermore, if participants all answer 50% of the questions correct in the no steering condition, they will all have the same perceived level of difficulty.

When setting up a noisy environment for participants with hearing impairments, it is usually a good idea to adjust the SNR to their hearing. The first idea was to use a Hearing In Noise Test (HINT) to adjust the ratio between noise and speech signal for every participant based on their Speech Reception Threshold (SRT), [Nilsson et al., 1994]. The HINT consists of stationary background noise and a speech signal, which consists of several different sentences. The participant then hears a speech signal in noise and repeats as much of the sentences possible. Through an adaptive staircase procedure, where the speech signal gets louder or quieter according to how well the participant answers, an SRT at 50% is collected. Every run consists of 20 sentences.

It is beneficial to run the test three times with three different groups of sentences to calculate a mean SRT and therefore have a better estimate of the participants' hearing threshold in noise.

Even though the HINT gives a good estimate of participants hearing in noise, it does not take into consideration that visual cues can help understand speech in noisy environments, which became clear during the pilot tests. Since a method to find each participant's specific SRT at 50% with audio-visual material was not available and the pilot tests showed the tasks to be too easy with participants' SRT found during HINT, it was decided to have a fixed SNR for every participant. According to Bernstein and Grant [2009], hearing impaired can understand 50% of speech at around -6 dB SNR when audio-visual material is also present. However, Bernstein and Grant [2009] did an experiment with short sentences of five words, while this master thesis investigates speech intelligibility while following a conversation and answering questions. Furthermore, Xia et al. [2017] investigated the understanding of speech when participants only had to identify the correct sound to make a correct response and when participants had to understand what they heard to make a correct response. From their experiment, Xia et al. [2017] concluded that participants were performing significantly better when just having to identify the correct sounds than understand the whole story. Because Bernstein and Grant [2009] is not measuring the same speech intelligibility, and since it is assumed that it will be harder to follow a whole conversation in -6 dB SNR, and therefore have a worse score than 50%, the SNR was set to -3 dB.
Even though the HINT could not be used to adjust the SNR, it was still measured during the experiment to understand the participants hearing in noise.

4.3 Conditions

To investigate the different hypotheses, it seems beneficial to let participants go through four different conditions to gather as much information as possible. A hard eye gaze steering and soft eye gaze steering condition are important to investigate how the eye gaze steering works and what the benefits are by using them. A no steering condition is included as a control condition to see how much eye gaze steering improves the tasks. A familiarization round is furthermore included for two purposes; to be sure the participants understand the tasks and see how people with hearing aids behave in calm surroundings. The different conditions will be explained further in the following.

Familiarization

Participants are introduced to a familiarization condition, where video clips with a multi-talker scenario and questions are presented. In this condition no background noise is present, the dialogue is presented at 62 dB and the monologue is furthermore reduced with 6 dB. This is done to have an easy start to the experiment, where the participants get familiarized with the procedure and get in the rhythm of answering the questions. Furthermore, the calm surroundings give an indication of how participants behave in an easier condition, which can be beneficial for later comparisons of conditions.

During the familiarization round participants are presented for 1 + 11 questions that they should answer. The first question is a try-out question (Q56 from appendix B) and is presented as the first question for every participant in the familiarization round. Data from this question is not used for further analysis. Additionally, participants are asked to answer the NASA-TLX questionnaire after this condition.

No steering

In the no steering condition, participants are presented with a multi-talker scenario in a noisy environment. Here background noise is present at 65 dB, and the monologue is played at the same level as the dialogue, 62 dB.

During the no steering condition participants are presented for 11 questions that they should answer. Additionally, participants are asked to answer the NASA-TLX questionnaire after this condition.

Hard eye gaze steering

In the hard eye gaze steering condition, participants are presented with a multi-talker scenario in a noisy environment. Background noise at 65 dB is present at all times, while the eye gaze of the participant is determining how much the talkers will be enhanced. The starting point for the speech signals is at 62 dB before any reduction or enhancement. During hard eye gaze steering the talker which the participant looks at, at every specific moment, is enhanced, while speech signals from other talkers are being reduced.

During the hard eye gaze steering condition participants are presented for 11 questions that they should answer. Additionally, participants are asked to answer the NASA-TLX questionnaire after this condition.

Soft eye gaze steering

In the soft eye gaze steering condition, participants are presented with a multi-talker scenario in a noisy environment. Background noise at 65 dB is present at all times, while the eye gaze of the participant is determining how much the talkers will be enhanced. The starting point for the speech signals is at 62 dB before any reduction or enhancement. During soft eye gaze steering, eye gaze is being measured over time, and enhancement is applied in the directions where the eye gaze is measured within the time period, according to how much they look in each direction. The talkers that are not looked upon during the time period is reduced.

During the soft eye gaze steering condition participants are presented for 11 questions that they should answer. Additionally, participants are asked to answer the NASA-TLX questionnaire after this condition.

4.4 Tasks for the participant

When participants entered the experiment, they first went through the HINT, explained in section 4.2, to gather insight about their hearing in noise. Here participants listened to difference sentences in noise and had to repeat as much as possible.

During the experiment participants went through the different conditions, saw 11-12 video clips per condition and answered the developed questions for each video clip. Before a video clip was shown, participants read the question out loud. This was done for the test facilitator to control that the video clip and questions matched, but also for the participant to have a better chance to understand and remember the question. Furthermore, if participants had any doubt related to the questions, the test facilitator could help them understand the question before showing the video clip.

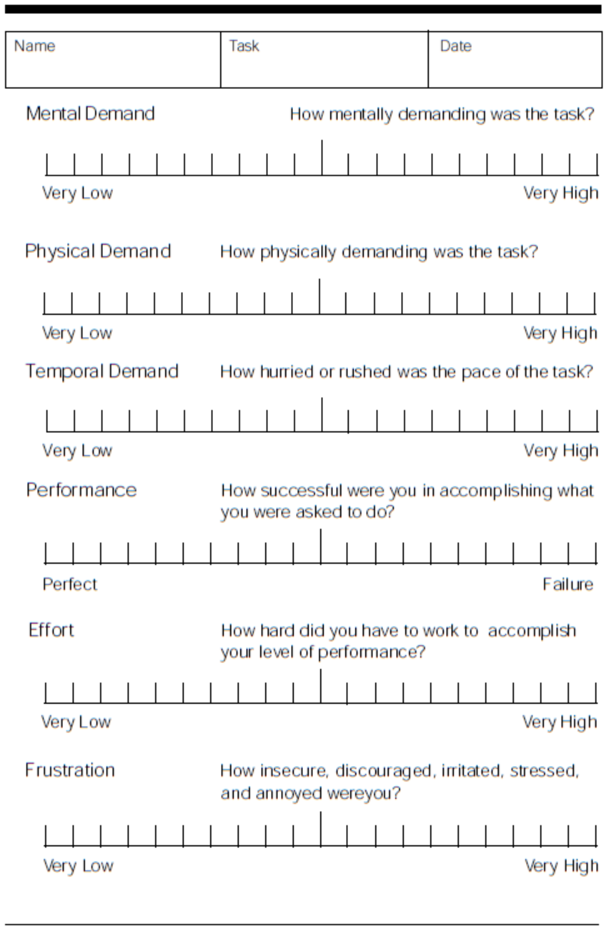
When the video clip was shown, participants were instructed to press a green button as soon as they knew the answer to the question. The green button was the 0-button on a keyboard with a green piece of tape over. When the button was pressed, the video shut down and participants gave their answer to the question orally to the test facilitator, who noted the answer and whether it was correct or wrong.

After all questions were answered during each condition, participants were asked to answer a NASA Task Load Index (TLX) questionnaire, [Human Performance Research Group and NASA Ames Research Center]. The questionnaire can be seen in figure 4.2 and is a widely used tool to measure workload. When participants answer the NASA-TLX it is possible to gather a subjective measurement of their workload, hereunder subjective measurements of the mental, physical, and temporal demand of the task, their own performance rating, how much effort they used, and how frustrated they were during the task.

To understand the importance of the different aspects of the TLX questionnaire, participants were, after all conditions were done, asked to do a weighting of the aspects. This was done by presenting one aspect across another and then have participants circle the one aspect they found the most important. In total, 15 circles were made by each participant and afterward used by the facilitator to weigh the different aspects according to how important each participant found them.

The questions in the NASA-TLX questionnaire is in English, while the rest of the test was in Danish. If participants do not understand English, this can be a problem.

However, it can also be problematic to translate a validated questionnaire without losing the intended meaning of the words. It was decided to hand out the questionnaire in English, but have a translation to give participants if they did not understand the questions. This translation was based on six English and Danish speaking persons, who translated the questionnaire separately. Since the translations were very much alike, see appendix C, this was assumed acceptable to present to participants, only if they had problems understanding the English words.



Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 4.2: The NASA Task Load Index (TLX) questionnaire that participants answered after every condition to measure the workload of the task, [Human Performance Research Group and NASA Ames Research Center].

After the four conditions and the weighting of the NASA-TLX aspects, an exit-interview was conducted. This was done to get insights and qualitative data from the participants to support the quantitative findings and answer hypothesis 3. It should be noted that participants at the beginning of the experiment knew that they would experience eye gaze steering, but not at which point they would experience which condition. During the exit-interview (after point two) they were told the difference between the conditions and how each condition had worked. The exit-interview had the following questions:

- How did it go? Did you experience any difficulties (fatigue, too much noise, having to force the eye gaze, close to giving up)?
- Did you notice a difference between the different conditions? Which differences?
- Did you find eye gaze steering helpful when trying to follow the conversation?
- Which of the two forms of eye gaze steering did you prefer?
- Which of the two forms of eye gaze steering felt most natural?
- Are there any situations where you would prefer one form of eye gaze steering over another? Which situations?
- Would you like for your hearing aid to be controlled in this way?
- What did you think about the questions presented to you?
  - Were they easy to remember?
  - Was it clear when the answer was given?
- Could you respond as quickly as possible?
- How did you like the actors?
- Do you have other comments?

4.5 Balance in the experiment

The developed questions have not been tested to see if they have the same level of difficulties, they have only been tested to see if they were understandable and possible to answer. To exclude bias from the same questions being asked in the same order or in the same conditions, everything in the experiment was balanced. This was done as described in the following.

In appendix B the different questions can be seen. The questions used are of different kinds, meaning that the answers are presented in different ways in the conversation. There are normal questions, where the keyword word from the question (e.g cat) is spoken before the answer (e.g brown). An example of this could be a question "What color is the cat?" and the answer within a sentence "The cat is brown". There are also questions, where the corresponding answer is spoken before the keyword from the question. An example of this could be the same question "What color is the cat?" and the answer within the sentence "There is a brown cat". Furthermore, there are questions that have the answer to be the first thing in the next talker's sentence. Questions were balanced to have different kinds of questions be equally divided between conditions. Additionally, the different videos were balanced out to be almost equally presented in each condition. This ended up with four balanced groupings of questions.

To not always present the questions in the same order under each grouping, the order of the questions is counterbalanced following a Latin Square counterbalancing, [Field and Hole, 2003, pp. 84-86].

Lastly, the conditions should also be balanced to avoid bias. The familiarization round was always kept as the first condition, while the no steering, hard eye gaze steering and soft eye gaze steering condition were completely counterbalanced.

In appendix A.2 the order of questions and conditions for each participant can be seen. During the experiment, if a participant was excluded for any reason, the same order of questions and conditions were repeated for another participant to maintain the overall balance of data.

4.6 Physical setup of experiment

For the experiment following equipment was used:

- Samsung 88” 4K curved TV screen to show the multi-talker scenarios in live sizes.
- Dikablis Eye Tracking Glasses with motion trackers placed on top to track the eye gaze direction.
- Four Vicon motion tracker cameras pointing towards the TV screen and the participant wearing the eye tracker.
- Computer with the Vicon 3.7 and D-lab version 3.45.
- MATLAB version R2016a with the developed program.
- Audiovisual material with a dialogue and a monologue.
- Clean speech signal from talkers in videos.
- Three loudspeakers to enhance the sound in the direction of the talker placed behind the screen at the position of the talker.
- Three loudspeakers below the TV screen to play noise under different conditions.
- Tracks with pedestrian noise to create a noisy environment.
- Table and chair for the participant, placed to approximately have the left talker at a 20 degree angle, the middle talker at a 0 degree angle, and the right talker at a -20 degree angle.
- An extra keyboard with a piece of green tape over the 0-button for participants to press whenever they knew the answer to the question.
- Printed questions easily displayed for the participant.

When the setup was complete, all speakers were calibrated with stationary noise. The setup can be seen on figure 4.3.

(a) The experiment setup with loudspeakers, TV screen with audio-visual material, keyboard with the green button, and questions.

(b) The experimental setup with loudspeakers, TV screen with audio-visual material, a test participant with the eye/head tracker on, and motion cameras tracking the eye/head tracker.

Figure 4.3: The experimental setup without and with a test participant.

4.7 Software

To run this kind of experiment, a program controlling and collecting data during the experiment was developed. The program was based on the flowchart shown in figure 4.4 and written by employees at Oticon and Eriksholm Research Centre. First of all, the videos and audio are loaded into MATLAB and synchronized. Afterward, different parameters as the form of the gain curve, the sound pressure level of the background noise, and the SNR to determine the sound pressure level of the speech signal are defined. Information about conditions and which video clip to play should be given to the program. This was done by developing a .csv file for every participant, where the trial number had information about which video to play, where to start and stop the video, and which condition it was. The .csv files for every participants can be seen under appendix A.2. During the experiment, the facilitator only typed in a trial number from 1 to 45, and the program did the rest. After the trial number is typed in, the program starts continuously measuring the eye gaze and saving this for every trial. Furthermore, the program is reading which condition to act upon.

As described in section 4.3, the different conditions have different enhancement of the background noise and speech signals. How much the talkers are enhanced or reduced during the two different eye gaze steering conditions are determined by where they look and by the gain curve parameters. The math behind both hard and soft eye gaze steering is described in detail in appendix D. What is important to note is that the hard eye gaze steering will apply the whole gain in the direction of the eye gaze, while the soft eye gaze steering will divide the gain to the different areas that are being looked at according to how long time the eye gaze was measured in the specific directions.

The eye gaze was measured with a Dikablis eye tracker, [Ergoneers], with motion trackers upon it that a Vicon motion tracker system, [Vicon], could pick up upon. With the Diklabis eye tracker, the movement and direction of the eyes were measured, while Vicon translated this to fit in the room. By using both Dikablis and Vicon it was possible to see in which direction within the room the participant was looking, and with a calibration of the room and the eye/head tracker, it was furthermore very precise. For the program to understand where the talkers were placed, motion trackers were set on top of the TV screen according to where the middle of the talker's head was placed. When the measured eye gaze aligned with the head of the talkers/the motion trackers on top of the TV screen, the eye gaze was understood as participants looking at a talker.

4.7.1 Parameters

In the program, different parameters could be changed, which had to be decided beforehand. This was the before-mentioned level of background noise and gain curves, but also the time eye gaze should be measured over in the soft eye gaze steering

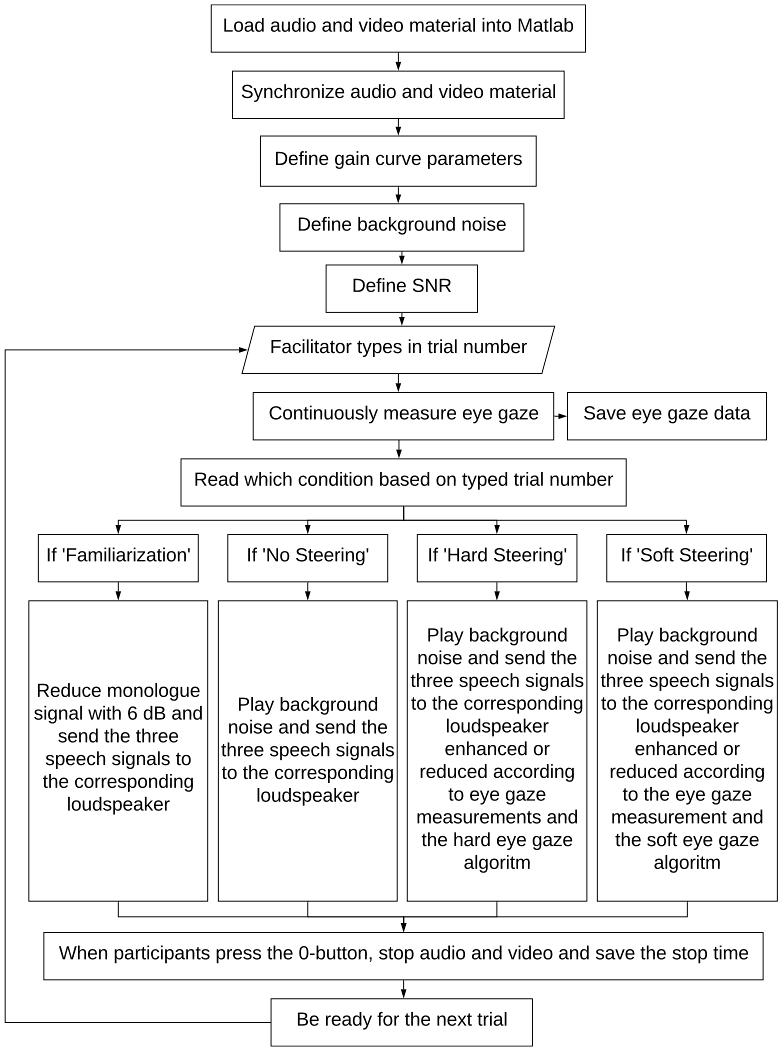


Figure 4.4: The flow diagram explaining what the program developed for the experiment does.

condition. What the different parameters should be set to, was decided in meetings between the master student and employees at Oticon and Eriksholm Research Centre.

Background noise

As described, the background noise was set to 65 dB at the listener's position and as the noise in the HINT kept constant. With the -3 dB SNR, speech signals were therefore presented at 62 dB as the starting level, and the monologue during the familiarization round was presented at 56 dB. This seemed to be a reasonable noise level that was both loud enough to hear for the participants and also loud enough to somewhat mask the talkers when they were not enhanced. Even though the background noise was loud enough to hear, it was not so loud it would be uncomfortable when the speech signal was enhanced to an SNR above 0 dB during the eye gaze steering conditions. The background noise throughout the experiment was not affected by the eye gaze steering.

Gain curves

The gain curves simulating a beamformer followed a Gaussian shape and had a range from -10 dB to +8 dB, which can be seen in figure 4.5. The blue gain curve is simulating a beamformer in the direction of the left talker, while the red and yellow is simulating a beamformer in the direction of the middle and right talker, respectively. This meant that when participants looked exactly where the talker was defined to be, an enhancement of 8 dB of the speech signal was made, and speech signals outside the gain curve were reduced by 10 dB. The reduction was chosen to have the effect of a big penalty of looking in the wrong direction, while the enhancement was chosen to be noticeable in both hard and soft eye gaze steering without being too loud for participant and facilitator in the room. The width of the gain curve was set to be wide enough to include small eye gaze errors or the talker moving their head a little from side to side. However, it was not set to be so wide that it would include enhancement of the other talkers, which can be seen in figure 4.5. In reality, hearing aids might not be able to enhance or reduce as much as described in all conditions. Furthermore, it might not be possible to have beamformers be as narrow as described. Parameters were chosen to be a bit extreme to clearly see the benefits or disadvantages of eye gaze steering.

Measurement time of eye gaze

When determining how much time the soft eye gaze steering should measure over, several suggestions were made. If the eye gaze was measured over a short time, e.g 2 seconds, the enhancement would quickly tune in on the new talker, but it would not be able to still enhance the last talker if a shift back and forth was made. On the other hand, measuring eye gaze over a long time, e.g 10 seconds or more, would make it easier for participants to shift between talkers and still hear the first words of the next talkers sentence, but if the eye gaze was kept for a long time on one talker and then shifted to the next, it would take a long time for the enhancement to reach maximum on the next talker.

According to Heldner and Edlund [2010] the maximum gaps between two talkers are at two seconds, which means it would be beneficial to have the measurements for more than two seconds. However, it is not possible to say for how long time a person will continue talking, which makes it hard to estimate for how long eye gaze should be measured over. Since no measurements of turn-taking or time talking on these specific videos made, and due to the video clips maximum being 30 seconds long, it was decided to measure the eye gaze over five seconds during soft eye gaze steering. This was done to experience a middle ground between too short and too long measurements, and because participants then had the time to experience the

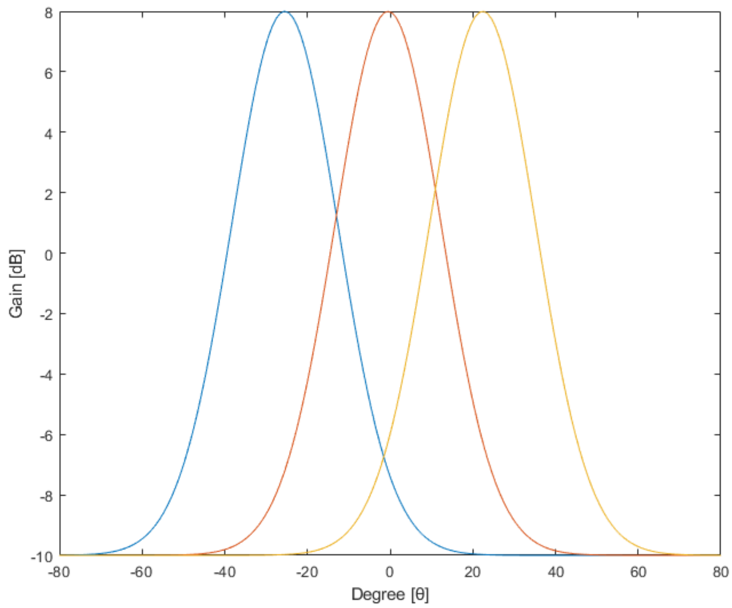


Figure 4.5: The gain curves used in the program. Here the maximum gain is at 8 dB and the minimum gain is at -10 dB. The blue gain curve is simulating a beamformer in the direction of the left talker, while the red and yellow is simulating a beamformer in the direction of the middle and right talker, respectively.

enhancement change within the 18-30 seconds video clip. Furthermore, five seconds seemed to generally fit the conversation in the videos, for the participant to have a chance of looking back and forth and hearing enhancements on both talkers in the dialogue.

4.8 Test participants

For the experiment, 15 people with slight to moderately severe hearing impairments were booked for the experiment. Audiograms from participants can be seen in appendix E. All participants were wearing their own hearing aid. The two first participants (two males, 40-53 years old) were used as pilots for testing out the setup with hearing-impaired people. Data from these will not be used since the sound pressure level of the background noise was changed after these two participants.

The remaining 13 participants can be seen in table 4.1. They were all socially active and moderate to very interested in new technology. All participants were connected to Eriksholm Research Centre, which means that they are being compensated with free hearing aids. All participants were very interested in participating in the experiment.

TP	Gender	Age	Hearing aid	Hearing aid experience
3	Female	21	CIC	7 months
4	Female	73	BTE	10 years
5	Male	81	BTE	>10 years
6	Male	71	BTE	>12 years
7	Male	70	BTE	10 years
8	Male	25	BTE	7 years
9	Male	21	BTE	Since childhood
10	Female	39	BTE	15 years
11	Male	77	BTE	7 years
12	Female	77	BTE	6 months
13	Female	76	BTE	11 years
14	Male	76	BTE	45 years
15	Female	79	BTE	10 years

Table 4.1: Table over participants (TP) used in the experiment.

4.9 Procedure

Throughout the experiment, the participants encountered the following:

1. Introduction to the experiment
2. Hearing in noise test (HINT)
3. Calibration of eye/head tracker
4. First condition which acts as a familiarization round
    (a) 1+11 video clips and questions
    (b) NASA-TLX questionnaire
5. Second condition with either eye gaze steering or no steering
    (a) 11 video clips and questions
    (b) NASA-TLX questionnaire
6. Break
7. Third condition with either eye gaze steering or no steering
    (a) 11 video clips and questions
    (b) NASA-TLX questionnaire
8. Fourth condition with either eye gaze steering or no steering
    (a) 11 video clips and questions
    (b) NASA-TLX questionnaire

9. Subjects weigh the aspects of NASA-TLX
10. Exit-interview
11. Thank you and goodbye

Participants participating in the experiment were booked for two hours, which included breaks along the way. When running the experiment, point 4, 5, 7, and 8 were balanced as described in section 4.5. The introduction to the experiment followed a manuscript, seen in appendix F.

Chapter 5

Results

From the experiment explained in the previous chapter, a lot of data was collected. To understand the user experience of the eye gaze steering and interpret which form of eye gaze steering, if any, helps the user the most, data is analyzed. The analysis is done by using Rstudio version 1.1.463, IMB SPSS Statistics version 25, and MATLAB 2019b. In the following chapter results from the analysis will be presented, which include analysis of speech intelligibility, workload, eye gaze data, and the qualitative data gathered from the exit-interview and observations.

Information about participants, answers to the questions, and answers to the exit-interview can be seen in appendix A.3, while the workload scores and percentage of correct answers can be seen in appendix A.4. SRTs measured from the HINT can be seen in table 5.1.

During the experiment, it was observed that the eye/head tracker and the program did not perform well for participant 8. For unknown reasons, the enhancement was shifted approximately 10 degrees to the right side, which resulted in the right talker being enhanced more than the middle talker when the eye gaze aligned with the middle talker, and the middle talker being enhanced more than the left talker when eye gaze aligned with the left talker. Due to this, data from participant 8 was not included in the data analysis. After this participant, the offset in the program was corrected, so the enhancement functioned as intended during the rest of the experiment.

TP	SRTs from HINT
3	4.35 dB
4	3.48 dB
5	4.59 dB
6	3.21 dB
7	4.63 dB
8	6.11 dB
9	5.73 dB
10	2.92 dB
11	4.71 dB
12	3.54 dB
13	3.90 dB
14	7.38 dB
15	3.60 dB

Table 5.1: Table over participants' (TP) SRT scores from the HINT.

5.1 Speech intelligibility

From the answers the participants provided for each question, the percentage of correct answers is calculated for each condition. The number of correct answers that were given during each condition is visualized in figure 5.1. On the boxplot, it is possible to see that the familiarization round provided the most correct answers, while no steering provided the least correct answers, which fits with the assumption that familiarization and no steering should be the easiest and hardest condition, respectively. From the boxplot, it is also possible to see that the eye gaze steering conditions lie in between familiarization and no steering, and that hard eye gaze steering seems to provide more correct answers than soft eye gaze steering when looking at the interquartile range of the boxplots.

To investigate if there is a difference in speech intelligibility between the different conditions, the conditions are analyzed using a repeated measures analysis of variance (ANOVA). The repeated measures ANOVA is used because data comes from a within-subject design and is collected in several conditions. Before running the analysis, assumptions like normality (if the data, in this case the percentage of correct answers, in each condition follows a normal distribution) and sphericity (whether the difference between conditions is similar) are checked, [Field et al., 2012, pp. 412 & 551]. In table 5.2 results from testing the assumptions can be seen. As shown, only one p-value is below 0.05, the p-value for normality during familiarization. This means that the data within the familiarization condition is not following a normal distribution. Besides this, the results show that data is not significantly different from a normal distribution and that the differences between conditions are similar. Assumptions for a repeated measures ANOVA is therefore partly met. Since [Field et al.,

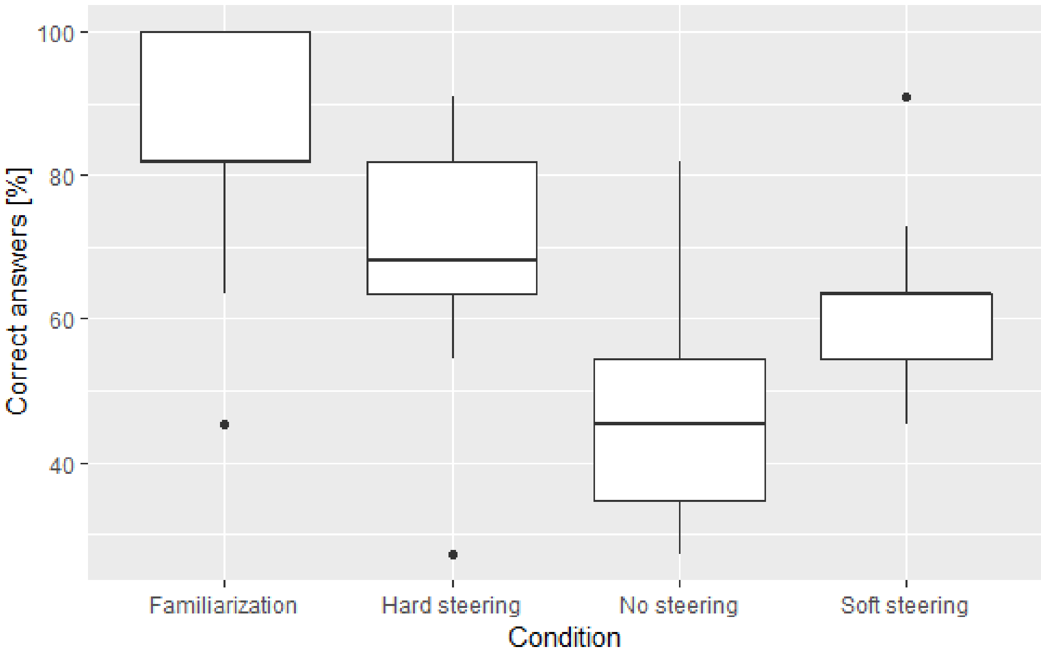


Figure 5.1: Boxplot showing the results from measuring speech intelligibility in every condition by how many questions participants could answer correctly in percentage.

2012, p. 413] argues that an ANOVA is rather robust towards violations of normality as long as the group sizes are similar (which they are in a within-subject design), it is decided to still run and trust the outcome of a repeated measures ANOVA analysis.

	Familiarization	No steering	Hard steering	Soft steering
Shapiro-Wilk normality test	W = 0.83 p = 0.023	W = 0.93 p = 0.371	W = 0.89 p = 0.121	W = 0.87 p = 0.058
Mauchly's test for sphericity	W = 0.454 p = 0.177			

Table 5.2: Table showing the results from running a Shapiro-Wilk test and a Mauchly's test to test for normality and sphericity, respectively. Assumptions are met for sphericity since the p-value does not show a significant difference. For normality, the assumptions are partly met since one condition shows a significant difference from a normal distribution.

A repeated measures ANOVA is made, and results can be seen in table 5.3. The results show a significant difference between conditions, since the $p < 0.05$. Furthermore, the effect size, generalized eta-squared (GES), is presented in the table. This effect size also shows a large effect, since $GES > 0.138$, [Phycho Hawks, 2010; Field et al., 2012]. To understand where these differences are, a post hos test is made with a Bonferroni correction. Results from the Bonferroni correction can be seen in table

5.4 and show a significant difference between all conditions, except hard and soft eye gaze steering. Furthermore, the effect size is also calculated for each comparison by using Cohen's d. Since Cohen's d is divided into small (0.2), medium (0.5), and large (0.8), [McLeod, 2019], it is possible to see that the effect size is large in between all conditions, except hard and soft eye gaze steering.

As seen, the results found from analyzing speech intelligibility data indicate a difference between no steering and eye gaze steering, which means that hypothesis 1(a) is accepted. However, since no significant difference is found between hard and soft eye gaze steering, and hard eye gaze steering seems to perform a little better, hypothesis 2(a) is not accepted.

Repeated-measures ANOVA	$F = 20.1$ $p = 1.342 \cdot 10^{-7}$
Generalized eta-squared	0.44

Table 5.3: Repeated-measures ANOVA showing a significant difference between conditions and generalized eta-squared showing a large effect size.

	Familiarization	Hard steering	No steering
Hard steering	$p = 0.00078$ Cohen's $d = 0.96$	-	-
No steering	$p = 0.00013$ Cohen's $d = 2.26$	$p = 0.01313$ Cohen's $d = 1.28$	-
Soft steering	$p = 0.01130$ Cohen's $d = 1.51$	$p = 1.00000$ Cohen's $d = 0.41$	$p = 0.02074$ Cohen's $d = 1.02$

Table 5.4: The table shows a Bonferroni correction of the repeated measures ANOVA. The p-values show a significant difference between all conditions, except hard and soft eye gaze steering. Furthermore, the table shows Cohen's d (the effect size), which is large between all conditions except hard and soft eye gaze steering.

5.1.1 The benefit of eye gaze steering in specific situations

In section 2.2.1 it is mentioned that a problem with hard eye gaze steering of beamformers can be that the first few words in the next talker's sentence is not heard if the next talker starts talking before the eye gaze is in this direction. Furthermore, the soft eye gaze steering might be helpful in this situation, if the hearing aid user is looking back and forth between the talkers. In appendix B all questions used in the experiment are presented, including the questions that have the answer in the video placed as the first thing in the next talker's sentence. To investigate if hard or soft eye gaze steering makes it harder or easier to answer these kinds of questions, the data from only these questions are analyzed. In figure 5.2 a boxplot over the percentage of correct answers given under each condition to these specific questions is presented.

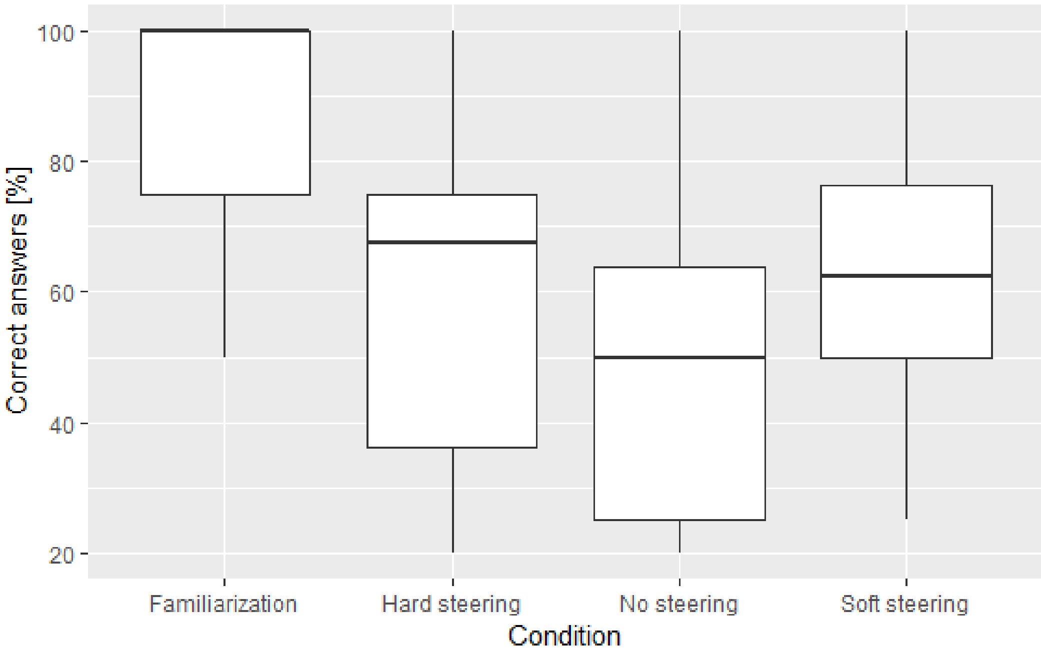


Figure 5.2: Boxplot showing the percentage of correct answers given to questions that have the answer at the beginning of a new talker's sentence for each condition.

On the boxplot, figure 5.2, it can be seen that questions under the familiarization round are still answered correctly most of the time, while questions under no, hard, and soft eye gaze steering are not answered correctly as many times. To investigate if there is any difference between the conditions, assumptions for a repeated-measures ANOVA are checked and a repeated measures ANOVA is conducted. Results from checking the assumptions can be seen in table 5.5. All assumptions are okay, except for familiarization, which still does not have normalized data. Since the assumptions are partly met, it is decided to trust the outcome of the repeated measures ANOVA, which can be seen in table 5.6.

	Familiarization	No steering	Hard steering	Soft steering
Shapiro-Wilk normality test	W = 0.73 p = 0.002	W = 0.89 p = 0.109	W = 0.91 p = 0.211	W = 0.93 p = 0.403
Mauchly's test for sphericity	W = 0.38 p = 0.098			

Table 5.5: Table showing the results from running a Shapiro-Wilk test and a Mauchly's test to test for normality and sphericity, respectively. Assumptions are met for sphericity since the p-value does not show a significant difference. For normality, the assumptions are partly met since one condition shows a significant difference from a normal distribution.

Repeated measures ANOVA	F = 6.67 $p = 1.206 \cdot 10^{-3}$
Generalized eta-squared	0.28

Table 5.6: Repeated-measures ANOVA showing a significant difference between conditions and generalized eta-squared showing a large effect size.

The results from the repeated measures ANOVA shows both a significant difference between conditions and large effect size. To understand where the differences are, a post hoc test with a Bonferroni correction is made. Results from the Bonferroni correction can be seen in table 5.7 and show a significant difference between the familiarization round and the hard and no steering condition, as well as a large effect size between the familiarization and the other three conditions. Besides this, no significant differences are detected. From the results, it is understood that hard or soft eye gaze steering does not make a difference in hearing the answer to a question when the answer is spoken as the first thing in the next talker's sentence.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.0016 Cohen's d = 1.15		
No steering	p = 0.0196 Cohen's d = 1.65	p = 1.0000 Cohen's d = 0.43	
Soft steering	p = 0.0672 Cohen's d = 1.14	p = 1.0000 Cohen's d = 0.14	p = 0.5756 Cohen's d = 0.60

Table 5.7: The table shows a Bonferroni correction of the repeated measures ANOVA. The p-values show a significant difference between the familiarization round and the no and hard eye gaze steering condition. Furthermore, the table shows Cohen's d (the effect size), which is large between the familiarization round and the other three conditions. No other differences are detected.

5.2 Workload

During the experiment, participants answered the NASA-TLX questionnaire after each condition. Furthermore, they also weighted the aspects according to which aspects they found most important in this scenario. From the questionnaire and weighting, a total workload score can be calculated for each condition and analyzed for any differences. Furthermore, it is also possible to look at each aspect individually to see if there is any difference between the conditions in any of the aspects.

5.2.1 Overall workload

As described in section 4.4 participants weighted the aspects of the NASA-TLX. In appendix A.4 and in table 5.8 it can be seen how much weighting each aspect got.

From the following formula the total workload for the condition can be calculated, [Human Performance Research Group and NASA Ames Research Center]:

$$Workload = \frac{\sum_{i=1}^{6} aspect_i \cdot weigth_i}{15} \tag{5.1}$$

Here the *aspect* is related to the score of the six questions in the questionnaire and the *weight* is related to the number of times this aspect was circled by the participant.

TP	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
3	4	1	2	4	3	1
4	2	1	4	5	0	3
5	1	0	2	5	3	4
6	3	3	1	3	2	3
7	3	2	4	5	0	1
9	3	1	3	2	3	3
10	4	0	1	5	2	3
11	5	3	0	4	2	1
12	3	0	5	3	1	3
13	2	2	4	5	1	1
14	4	2	1	3	0	5
15	4	1	5	3	2	0

Table 5.8: Table showing how participants weighted the aspects of the NASA-TLX questionnaire.

On figure 5.3 a boxplot over the total workload for each condition can be seen. Here the familiarization round scores a lower workload than the no steering and both eye gaze steering conditions. Furthermore, there does not seem to be a difference between the no steering condition and both eye gaze steering conditions, since the interquartile range is overlapping between the boxplots.

To investigate if there is a difference in workload between the different conditions, the conditions are again analyzed using a repeated measures ANOVA. The assumptions are checked and results can be seen in table 5.9. As shown, p-values > 0.05 for normality, which means that this assumption is met. However, the assumption about sphericity is violated since $p < 0.05$. Violation of sphericity means that the repeated measures ANOVA might not be reliable, [Field et al., 2012, pp. 552-553]. Furthermore, the violation of sphericity can cause complications for the following post hoc test.

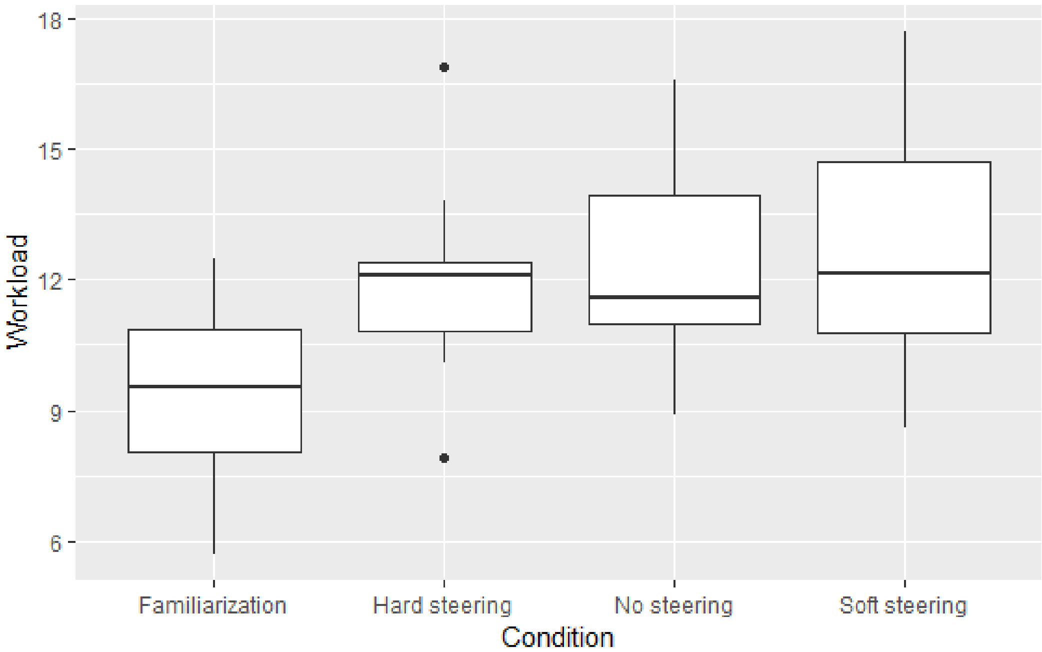


Figure 5.3: Boxplot showing the results from measuring workload in every condition by using the NASA-TLX questionnaire and weighting.

	Familiarization	No steering	Hard steering	Soft steering
Shapiro-Wilk normality test	W = 0.96 p = 0.837	W = 0.92 p = 0.283	W = 0.93 p = 0.375	W = 0.93 p = 0.430
Mauchly's test for sphericity	W = 0.17 p = 0.004			

Table 5.9: Table showing the results from running a Shapiro-Wilk test and a Mauchly's test to test for normality and sphericity, respectively. The assumption is met for normality since none of the p-values show a significant difference. For sphericity, the assumption is not met since the p-value shows a significant difference.

Even though one of the assumptions is violated, a repeated measures ANOVA is still conducted, but with the violation in mind. The p-value and effect size (GES) can be seen in table 5.10, which shows both a significant difference ($p < 0.05$) and a large effect ($GES > 0.138$). [Field et al., 2012, p. 553] mentions that when sphericity is violated, a post hoc test with a Bonferroni correction seems to be the most robust test in avoidance of type I errors (wrong rejection of the null-hypothesis). On table 5.11 results from a Bonferroni correction can be seen together with the calculated effect sizes. The results show a significant difference between the familiarization and the three other conditions. The effect sizes for familiarization compared with no, hard, and soft eye gaze steering also shows a large effect. For the other comparisons, no effects are detected. Looking at the boxplot on figure 5.3, these effects are also visualized, since the familiarization is not scoring as high in workload as the other three conditions.

As seen, the results found from analyzing the workload data does not indicate a difference between no steering and eye gaze steering, which means that hypothesis 1(b) can not be accepted. Furthermore, since no significant difference is found between hard and soft eye gaze steering, hypothesis 2(b) is not accepted either.

Repeated-measures ANOVA	F = 6.02 $p = 2.172 \cdot 10^{-3}$
Generalized eta-squared	0.23

Table 5.10: Repeated-measures ANOVA showing a significant difference between conditions and generalized eta-squared showing a large effect size.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.015 Cohen's d = 1.15	-	-
No steering	p = 0.067 Cohen's d = 1.24	p = 1.000 Cohen's d = 0.18	-
Soft steering	p = 0.073 Cohen's d = 1.29	p = 1.000 Cohen's d = 0.32	p = 1.000 Cohen's d = 0.14

Table 5.11: The table shows a Bonferroni correction of the repeated measures ANOVA. The p-values show a significant difference between familiarization and the other three conditions. No difference between the no, hard, and soft eye gaze steering is detected. Furthermore, the table shows Cohen's d (the effect size), which is large between familiarization and the other three conditions and small between the no, hard, and soft eye gaze steering condition.

5.2.2 Aspects of NASA-TLX

To look at each different aspect or question in the questionnaire, a boxplot is first set up, see figure 5.4. Here the different aspects under each condition can be compared. From the boxplot, it seems there might be a difference between mental demand in familiarization and the other conditions, but no other effects seem obvious. To gain a better overview and investigate any possible differences between the aspects, a one-way repeated measures multivariate analysis of variance (MANOVA) is conducted.
Before running a one-way repeated measures MANOVA, assumptions are checked. When doing this MANOVA it is assumed that data is measured on an interval or ratio scale, that the independent variable should consist of two or more categorical, related groups, and that there is sufficient data, [Laerd Statistics, 2018]. All of these assumptions are met. Furthermore, it is assumed that there are no outliers in the data, that there is multivariate normality (normality of each of the dependent variables for each of the related conditions), that there is a linear relationship between

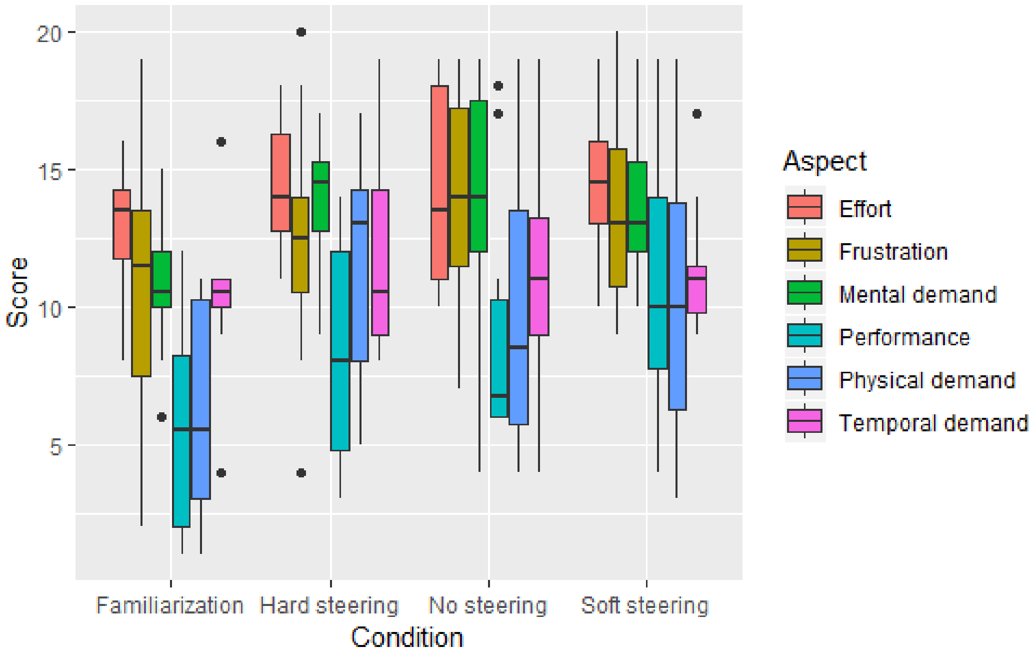


Figure 5.4: Boxplot showing the results for every aspect of the NASA-TLX questionnaire in every condition.

each pair of the dependent variables for each condition, and that the dependent variables are only moderately correlated. It is assumed that there is no significant correlation between the dependent variables since the dependent variables come from questions in a validated questionnaire. From the boxplot on figure 5.4 it is seen that there are outliers in the data. In appendix G a Shapiro-Wilk test for normality can be seen as well as scatterplots checking the linear relationship. Data seems to generally be normally distributed, except the temporal demand aspect, where the assumption is violated. Looking at the scatterplot matrices in appendix G, it is seen that the assumption about a linear relationship between each pair of the dependent variables is generally violated.

Two out of seven assumptions for a one-way repeated measures MANOVA are violated, while five of them are okay. The MANOVA analysis is still conducted, but will be interpreted carefully since not all assumptions are met. Results from this MANOVA can be seen in table 5.12. Since $p > 0.05$ there is no sign of a significant difference between the aspects in the different conditions. However, since not all assumptions checked out, the p-value is relatively small, and the effect size, partial eta-squared, is large, it is interesting to look at univariate tests, meaning the repeated measures ANOVA for each aspect.

One-way repeated measures MANOVA	F = 1.52 p = 0.107
Partial eta-squared	0.24

Table 5.12: The table shows results from a one-way repeated measures MANOVA and the effect size (Partial eta-squared). Even though no significant difference is detected, the effect size is large.

Before running a repeated measures ANOVA on each aspect, assumptions about sphericity is furthermore checked. The results from a Mauchly's test of sphericity can be seen in appendix G and show sphericity for all aspects, except temporal demand. On table 5.13 results from running a repeated measures ANOVA on each aspect of the NASA-TLX questionnaire can be seen. These results are with sphericity assumed, which is why temporal demand might not be trustworthy. The one-way repeated measures MANOVA showed no significant difference, but when looking at each repeated measures ANOVA, a significant difference is seen in mental demand, physical demand, performance, and effort, all with effect sizes that are interpreted as large. Since the boxplot, figure 5.4, did not show all these effects, they seem a little strange. The more conservative correction of a post hoc test, Bonferroni correction, is made to see if there is a difference between the conditions under the four aspects showing a significant difference.

	Repeated measures ANOVA	Partial eta-squared
Mental demand	F = 6.28 p = 0.002	0.36
Physical demand	F = 4.37 p = 0.011	0.28
Temporal demand	F = 0.48 p = 0.698	0.04
Performance	F = 3.71 p = 0.021	0.25
Effort	F = 3.00 p = 0.045	0.21
Frustration	F = 2.54 p = 0.073	0.19

Table 5.13: The table shows results from running several repeated measures ANOVA analysis - one on each aspect from the NASA-TLX questionnaire.

Table 5.14, 5.15, 5.16, and 5.17 shows the results from a post hoc test with a Bonferroni correction on the four aspects which showed a difference in the repeated measures ANOVA tests. Furthermore, the effect size found with Cohen's d is also presented in the tables. It is seen that the performance and effort-aspect do not show any difference between the conditions. Physical demand shows a difference between familiarization and hard eye gaze steering, while mental demand shows a difference between the familiarization round and the other three conditions, which was also seen in figure 5.4.

From the results found when investigating the different aspects in the NASA-TLX questionnaire, it is seen that the biggest difference found is under mental demand, where participants find it significantly more mentally demanding to do the task during no, hard, or soft eye gaze steering compared to the familiarization round.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.011 Cohen's d = 1.48	-	-
No steering	p = 0.063 Cohen's d = 0.95	p = 1.000 Cohen's d = 0.02	-
Soft steering	p = 0.009 Cohen's d = 1.22	p = 1.000 Cohen's d = 0.10	p = 1.000 Cohen's d = 0.08

Table 5.14: Mental demand: The table shows a Bonferroni correction of the repeated measures ANOVA made for mental demand. The p-values show a significant difference between familiarization and the other three conditions. No difference between the no, hard, and soft eye gaze steering is detected. Furthermore, the table shows Cohen's d (the effect size), which is large between familiarization and the other three conditions and small between the no, hard, and soft eye gaze steering condition.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.024 Cohen's d = 1.34	-	-
No steering	p = 0.198 Cohen's d = 0.80	p = 1.000 Cohen's d = 0.31	-
Soft steering	p = 0.218 Cohen's d = 0.84	p = 1.000 Cohen's d = 0.30	p = 1.000 Cohen's d = 0.02

Table 5.15: Physical demand: The table shows a Bonferroni correction of the repeated measures ANOVA made for physical demand. The p-values show a significant difference between familiarization and hard eye gaze steering. No difference between the other conditions is detected. Furthermore, the table shows Cohen's d (the effect size), which is large between familiarization and the other three conditions and small between the no, hard, and soft eye gaze steering condition.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.720 Cohen's d = 0.67	-	-
No steering	p = 0.100 Cohen's d = 0.83	p = 1.000 Cohen's d = 0.19	-
Soft steering	p = 0.130 Cohen's d = 1.22	p = 0.780 Cohen's d = 0.60	p = 1.000 Cohen's d = 0.41

Table 5.16: Performance: The table shows a Bonferroni correction of the repeated measures ANOVA made for performance. The p-values show that no significant differences between any conditions are detected. Furthermore, the table shows Cohen's d (the effect size), which is large between familiarization and the no and soft eye gaze steering condition and small or medium between all other conditions.

	Familiarization	Hard steering	No steering
Hard steering	p = 0.220 Cohen's d = 0.53	-	-
No steering	p = 0.700 Cohen's d = 0.42	p = 1.000 Cohen's d = 0.02	-
Soft steering	p = 0.220 Cohen's d = 0.60	p = 1.000 Cohen's d = 0.14	p = 1.000 Cohen's d = 0.10

Table 5.17: Effort: The table shows a Bonferroni correction of the repeated measures ANOVA made for effort. The p-values show that no significant differences between any conditions are detected. Furthermore, the table shows Cohen's d (the effect size), which is small or medium between all conditions.

5.3 Testing for order effects

During the exit-interview, three participants mentioned that they had to concentrate a lot during the experiment, but only one of the participants commented on being exhausted at the end of the experiment. To be sure that the results from measuring speech intelligibility and workload depended on the different conditions and not on how exhausted they got during the experiment, results were grouped into first, second, and last, depending on when in the experiment these results were obtained. In this part, the familiarization round is not taken into account since this condition was always the first one. First, second, and last is therefore equivalent to the second, third, and fourth condition that participants experienced during the experiment, respectively. Results can be seen on figure 5.5 and 5.6. From these boxplots, it seems that there is no difference between when the answers were given, which supports the participants' statements about not being exhausted during the experiment.

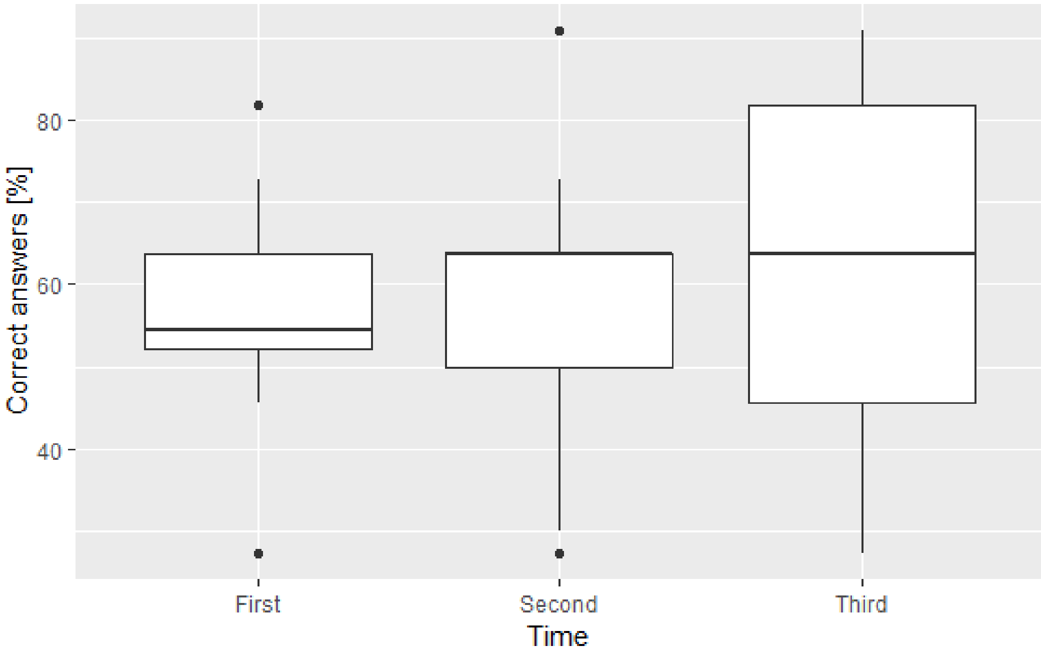


Figure 5.5: Boxplot showing the percentage of correct answers given at the first, second, and third condition, when the familiarization round is not taken into account.

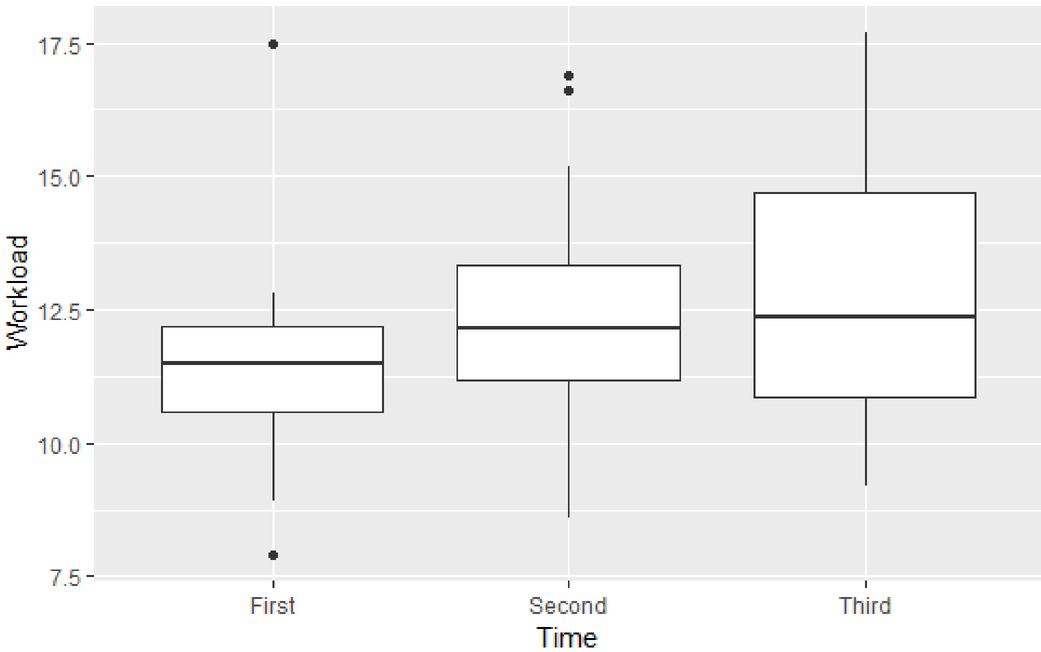


Figure 5.6: Boxplot showing the workload score given at the first, second, and third condition, when the familiarization round is not taken into account.

5.4 Eye gaze

During the experiment, eye gaze data was measured for each trial, which can be seen in appendix A.5. This eye gaze data makes it possible to show histograms over eye gaze during each condition and each video. However, due to the correction of the offset after participant 8, raw eye gaze data from participant 9 to 15 are not aligned with the -20, 0, and 20 degrees where the talkers were placed. The offset was approximately 10 degrees for all participants. Therefore the data from participant 9 to 15 was corrected 10 degrees back to match the data collected up until participant 8. This correction might not be perfect, which is why the shown histograms are only used to get an idea of participants' eye gaze direction during the experiment.

During the experiment four different videos were shown, called video 5, 9, 10, and 11. Between these videos, the talkers are not sitting exactly at the same spot, and the monologue and dialogue are switched around between the videos (in video 9 the monologue is to the left, while in the other videos the monologue is to the right). It is therefore beneficial to divide the videos when looking at eye gaze data under each condition. Histograms with eye gaze data over the four different conditions during video 10 is shown on figure 5.7, 5.8, 5.9, and 5.10. Histograms from the other videos can be seen in appendix H.

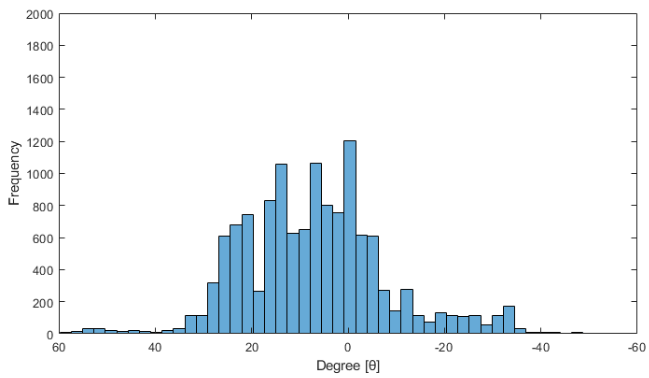


Figure 5.7: The histogram shows eye gaze data from the participants during video 10 under the familiarization round. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

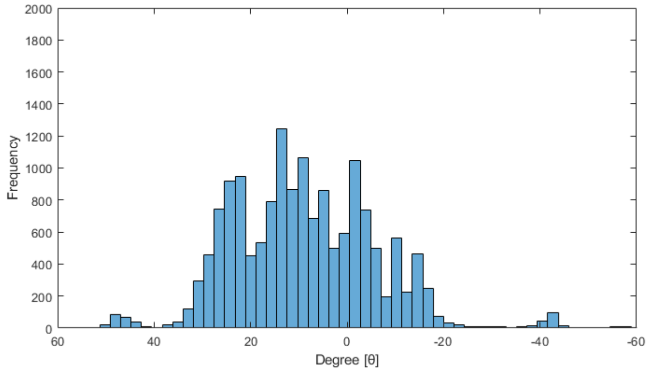


Figure 5.8: The histogram shows eye gaze data from the participants during video 10 under the no steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

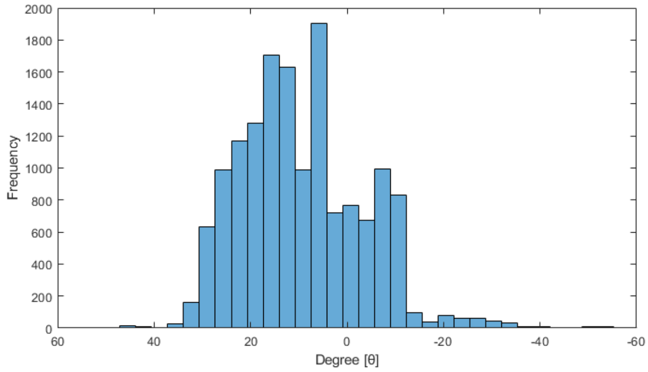


Figure 5.9: The histogram shows eye gaze data from the participants during video 10 under the hard eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

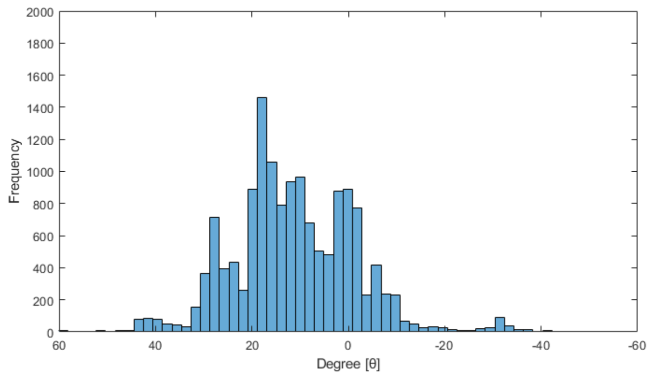


Figure 5.10: The histogram shows eye gaze data from the participants during video 10 under the soft eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

As seen in figure 5.7, 5.8, 5.9, and 5.10 all the different histograms, eye gaze behavior does not seem to change drastically between the different conditions. However, the histograms suggest that there might be a tendency of participants rarely looking at the monologue, while the talkers in the dialogue and the space around the dialogue are looked at a lot. This is supported by the participants who lip read, who will naturally look in the direction of the monologue, but was also the general observation during the experiment; participants would look in the direction of the dialogue because it is where they would find the answer to the question.

5.5 Qualitative data

During the experiment, qualitative data was gathered. This was done by having an exit-interview at the end of the test, where participants could give a qualitative response on which condition or form of eye gaze steering they liked the best. Furthermore, qualitative data was gathered during the experiment by observing and writing down notes, which can be seen in appendix A.3.

5.5.1 Eye gaze steering

Before being asked about the eye gaze steering, participants were asked if they noticed any difference between the conditions. Here four participants commented that the first condition, the familiarization round, was the easiest, and five participants commented that it only got harder throughout the test. Three participants commented that some of the talker's voices sometimes were clearer than others, and one commented that the monologue was sometimes reduced, but they did not seem to know why. Four participants commented on the background noise instead of the voice of the talker - that the background noise seemed to change.

When asked about if they found eye gaze steering helpful, only two participants had realized that the talker got enhanced when they looked that direction during the eye gaze steering conditions. Of the remaining 10 participants, three noticed something, but did not understand what they experienced, one noticed the monologue got less annoying, and one noticed the enhancement was slower during soft eye gaze steering. Since most of the participants did not exactly understand or notice the eye gaze steering during the experiment, and since they did not know how the experiment investigated eye gaze steering, both hard and soft eye gaze steering was explained to them before asking them what they preferred or what they thought they would prefer.

Hard eye gaze steering vs soft eye gaze steering

During the exit-interview, participants were asked which of the two forms of eye gaze steering they preferred and which they found most natural. Since most of them did not notice the eye gaze steering, these questions were also asked in a more general manner.

Out of 12 participants, 10 of them preferred the hard eye gaze steering. This was because they felt more in control and could hear the most with this form of eye gaze steering, which they felt gave them the best help to their hearing problem. Two participants liked the soft eye gaze steering better, either because they felt it would be better or because it then meant that their eyes should not be jumping around from talker to talker very fast all the time.

When asked about which form of eye gaze steering they found most natural, only four participants found the hard eye gaze steering most natural, while four participants found the soft eye gaze steering most natural. The remaining four thought it felt equally natural or they did not know what to answer since they did not feel a difference when trying it. The arguments for hard eye gaze steering being most natural were that they could hear, that they only want to hear what they are looking at, and that this form of eye gaze steering could help their hearing impairment the most. The arguments for soft eye gaze steering feeling most natural were that you could be interested in several things within your eyesight and that it would be confusing to change eye gaze all the time as with the hard eye gaze steering.

Even though a lot of participants were in favor of the hard eye gaze steering, they could also see some situations where they would prefer soft eye gaze steering. This was for instance if there were no distracting people talking or when following several people's conversation in a bigger assembly. However, the hard gaze eye steering would be preferred in bigger gatherings where they would wish to listen to one person. This could, for instance, be in a restaurant, at a theater, or at a cocktail party, where there could be a competing talker right next to them that they are not interested in listening to.

Participants were furthermore asked if they would be interested in having eye gaze steering in their hearing aids. Eleven participants said yes immediately, while one would like to try it in the hearing aid first. One participant mentioned in addition that it would be preferable to have both forms of steering to shift between and one mentioned that it would be nice to be able to turn eye gaze steering off when being in situations where information from all sides could be useful, such as taking the train. One participant actually had a pair of directional hearing aids beforehand, which brought a lot of joy. For this participant, it would be amazing to get something similar again.

As described, the results found from the exit-interview, regarding which form of eye gaze steering participants preferred, indicated that most participants generally prefer the hard eye gaze steering. Even though there are some situations where soft eye gaze steering is preferred, hypothesis 3 can not be accepted.

5.5.2 The audio-visual material and developed questions

During the exit-interview participants were asked what they thought about the questions presented to them, if they found the answers clear, and how they liked the actors/talkers in the videos.

Questions

Nine out of 12 participants commented that the questions are easy to understand and remember, while one participant commented that they were not always logical. Sometimes participants needed to look down and read the question again during the video clip. Four participants commented that this was the case, especially when the questions were long. By observing participants, it was clear that some questions were harder to answer than others. This was especially Q1, Q34, and sometimes Q59.

When asked about the answers, six participants commented that when they could hear, it was always clear what the answer was, while three participants commented that they sometimes were unsure if the question had been answered. This unsureness comes from the actors discussing back and forth before settling on an answer, which for instance was the case in Q7 and Q15. Two participants commented that they always knew when the answer was spoken in the dialogue, even though they did not always hear the answer. This was also seen during the observation, where participants sometimes looked like they were about to press the button, but did not, because they could not hear the answer-word even though they knew that the actors were talking about the related question.

During the experiment, it was also observed that when the question had "him" or "her" related to the talker who answered the question, participants looked mostly in that direction. This sometimes resulted in missing context, which made them unable to answer the question. Furthermore, some of the video clips overlapped in such a way, that an answer to another question was said during the video clip. However, none of the participants seemed to remember the answer, when they were asked the relating question.

Actors

When asked about the actors, 10 participants commented that they are good and seem to enjoy the job. Participants did have some trouble understanding the blond-haired guy, see figure 4.1, especially in the video clips where his voice was recovered by declipping and A-weighting. Five participants commented that he mumbles a lot and is hard to understand, and two of these actually gave up listening to him unless the question stated that he had the answer. Furthermore, four participants commented that the voices of the women were easier to hear and understand, especially the woman in the blue shirt, see figure 4.1.

5.5.3 Other observations

During the exit-interview, participants were asked how the experiment went, and if they experienced any complications. Here five participants commented that the experiment was fun and exciting. However, nine participants commented that it was usually hard to hear and understand the talkers, mainly when the background noise was present, but also if the audio and video were not synchronized, which happened a few times during the experiment due to MATLAB being overloaded. When the synchronization did not match, participants were not able to also lip-read, which two of them depended on. One participant also commented that it was hard to hear the talkers when they were reduced in volume, due to her looking the other way.

As described in section 5.3, participants do not seem to get exhausted during the experiment. Furthermore, none of the participants commented on being close to giving up or had to force their eyes somewhere. Two participants only commented that they might be exhausted if they had to listen in noise like this all day. Even though they did not feel like forcing their eye gaze somewhere, four participants had comments about looking in a specific direction:

- Participant 6: During the experiment he learned to look at the person talking to understand the most.
- Participant 11: During the familiarization he found a rhythm, where he could understand the talker he looked at the best.
- Participant 13: He turned his eye gaze to see the talkers and hear them better.
- Participant 14: Since he knew about the eye gaze steering concept beforehand he considered if it might be better that he looked between talkers to hear them both.

Even though most participants commented that they did not notice the eye gaze steering, these comments about eye movement can indicate that they felt a positive effect from the eye gaze steering after all.

Chapter 6

Discussion

To understand why some tendencies are found and how the experiment could have been done differently, the project and results are discussed. In this chapter, a discussion of the speech intelligibility, workload, sources of errors, and eye gaze steering in hearing aids are presented.

6.1 Speech intelligibility

As described in section 4.1.1 measuring speech intelligibility is often done by calculating the percentage of correctly identified words, while speech intelligibility in this master thesis is referring to the percentage of correct answers given to the questions. It is assumed that it is easier to repeat a few words from a sentence presented than to follow a conversation and figuring out the answer to a question, even though it is easier to get a lower score if just a single word is missed in the first-mentioned task. This is assumed both because there are a lot more words involved in a conversation and because the participant both have to remember the question and search for the answer. Therefore it might be hard to compare speech intelligibility in this experiment to speech intelligibility in other experiments. However, measuring speech intelligibility, or how well participants could follow the conversation, during the different conditions in this manner seems to be a very ecological way of doing it. Since the participants had to follow a real conversation, it gives a good indication of how eye gaze steering would actually work in reality, where users would also have to follow real conversations.

Even though the way of measuring speech intelligibility is different from how, for instance, Harrison [2018] did it, results still show a clear indication of eye gaze steering being better than no steering, which supports hypothesis 1(a). Results also show a clear indication of it being easier to follow the conversation and answer the questions with calmer surroundings as in the familiarization round. This is supported by comments during the exit-interview, were most people commented on the familiarization being the easiest.

Besides eye gaze steering being better than no steering when measuring speech intelligibility, it was also hypothesized that soft eye gaze steering would better than hard eye gaze steering. This hypothesis 2(a) is rejected since the results showed no significant difference or large effect between the two forms of eye gaze steering. Furthermore, when looking at the median of the boxplots in figure 5.1, hard eye gaze steering seems to get more correct answers than soft eye gaze steering, which is pointing in the opposite direction of the hypothesis.

In figure 5.1 a few outliers can be seen. These outliers come from participant 14, who had a high SRT and therefore had a harder time following the conversations, and participant 10, who had a lower SRT and found it quite easy to follow the conversation. If the background noise was adjusted to each participant as first planned, maybe these participants would not be outliers, and maybe a significant difference could then be seen between hard and soft eye gaze steering.

6.1.1 The benefit of eye gaze steering

During the experiment, it was noted that some participants commented on the talker being very clear or not clear at all. These comments were often related to the eye gaze steering conditions, where participants found the talkers clear whenever they looked at them or very hard to understand, when they looked at the other talker or down at the question. Furthermore, it was observed that when they could not answer the question during an eye gaze steering condition, it was often because they looked away. So even though participants did not all seem to notice the eye gaze steering, their comments were very much related to them experiencing the eye gaze steering conditions.

When analyzing the results it became clear that the time measuring eye gaze during the soft eye gaze steering might have been either too short or too long to actually be helpful for participants. It was either not long enough for participants to look back and forth between talkers and still have both talkers enhanced, or too long, so it took too much time for the next talker to be enhanced.

This was also seen when looking at the questions which had the answer as the first part of the next talkers' sentence. It was assumed that soft eye gaze steering would help hear the whole sentence, but the results fail to detect any difference in how many correct answers participants got between no, hard, and soft eye gaze steering. The lack of difference is probably due to 1) hard eye gaze steering is not giving people the opportunity to hear the first words of the new talkers' sentence, if they moved their eye gaze too slow or 2) soft eye gaze steering is taking too long to enhance the next talker, or not having measured eye gaze for long enough that the two talkers were both enhanced.

6.2 Workload

As seen in the results, the overall workload for no, hard, and soft eye gaze steering was significantly greater than for the familiarization, which aligns with participants commenting that the task got harder after the familiarization round. Furthermore, there does not seem to be a difference in the overall workload when comparing the no, hard, and soft eye gaze steering condition, which is supported by participants not really noticing that much of a difference between the last three conditions.

This lack of detected differences between three of the conditions can maybe come from the NASA-TLX questionnaire not being sensitive enough to show a difference between the conditions or maybe it stems from there not being participants enough in the experiment to see a difference. Furthermore, participants commented on the weighting of the aspects being hard to do. They did not all understand how to do it and found the task rather difficult. If they did not understand the task and therefore did not do the weighting correctly, then some aspects might have gotten more or less weight, which could pull the overall workload in the wrong direction. This could, for instance, be if an aspect was rated significantly different in the four conditions, but the weight on this aspect was too low to have the difference make an impact. However, when looking at the individual aspects from the NASA-TLX questionnaire, there do not seem to be many significant differences between the conditions. The only visible differences are between familiarization and the other three conditions in mental demand, and between familiarization and hard eye gaze steering in physical demand. Some of the other comparisons between familiarization and the other conditions also have large effect sizes, but similar to the overall workload, familiarization seems to score lower in the individual aspects. So even though the aspects might have been hard to weight, it is clear that the no, hard, and soft eye gaze steering conditions are not significantly different when measuring workload.

The high workload and the non-difference probably come from no, hard, and soft eye gaze steering being under highly demanding surroundings. Even though hard and soft eye gaze steering were supposed to help, eye gaze steering is a very new feature that participants only experienced for a short period of time. Since the two forms of eye gaze steering are new features, they can also be demanding to get to know, which can be what influences the results.

With more experience and longer trial periods of the concept, maybe the user will adapt to hard and soft eye gaze steering, which will make eye gaze steering easier to use and lower the workload. However, for this experiment, it was not shown to be easier to use either no, hard, or soft eye gaze steering when measuring workload, which is why hypotheses 1(b) and 2(b) are rejected.

6.3 Errors due to the experimental setup

In every experiment, different sources of error can occur. Since this experiment was not a completely ecological study and the experiment only tested the concept of eye gaze steering, errors could also occur due to the way the experiment was set up. In the following, the different parts of the experiment will be discussed with sources of errors as the focus.

6.3.1 Audiovisual material

During the experiment, participants followed a conversation between talkers on a big TV-screen. Since the talkers were showed on a screen instead of being present in the room, participants probably acted a little different than they would in real life. This could, for instance, be in the way participants could stare at the talkers without it being impolite or uncomfortable, even though this is a prerequisite for hard eye gaze steering to work. From the results, it is seen that participants look mostly in the direction of the dialogue, and not very much at the monologue. This could be due to the instructions, where they were told that the answer to the question would be in the dialogue. Since the goal for participants was to answer the question correctly, they directed their attention towards the dialogue. If present talkers were used in the experiment, the results might have been different - maybe participants would have looked more around the room, which could have made hard eye gaze steering less beneficial and more confusing. However, since some of the participants mentioned that they used lip-reading a lot and since Vatikiotis-Bateson et al. [1998], Buchan et al. [2008], Vatikiotis-Bateson et al. [1994], and Vertegaal et al. [2001] concluded that people tend to look at the person talking in noisy environments, it might also have been participants' natural behavior that was observed.

For the purpose of the experiment the audiovisual material seemed to work very well. Even though the situation was not completely ecological, participants still got a lot of visual cues from the talkers that they would also get in reality. This was as mentioned lip-reading, which was a tool that some participants were using in their everyday life. A few times it was observed that the audio and video did not start synchronized, which made it harder for those participants who used lip-reading. Luckily this glitch was only seen very few times during an experiment, but it could still be a source of error.

Another problem with the audiovisual material was the speech signal from the middle talker in video 9, who had his voice clipped during the recording. Even though several attempts were done to save the speech signal, participants still commented on this actor being hard to understand, since he mumbled a lot. Since he was so hard to understand, answers that he provided were sometimes very hard to hear for participants and participants often got annoyed with him, which could have influenced how well they answered questions. In general, the actors could influence how easy or hard it was to answer the questions since their way of articulating or gesturing could have made it easier or harder to hear and understand the answers. For future experiments, video 9 should probably be discarded. Furthermore, it could be beneficial to have videos with actors acting in the same way, so their voices or gestures did not influence the results.

Even though the results might be influenced by the actors in the videos, participants commented that the videos reflected the reality in a good way, since people also act differently and can be easy or hard to understand.

6.3.2 Speech intelligibility questions

The questions developed from the audiovisual material also generally seemed to work for the purpose of this experiment. If the questions were short, they were easy to remember and participants kept their eye gaze on the TV screen. If the questions were longer, participants found it beneficial to be able to look down and read the question again, even though this could have made the hard eye gaze steering condition harder. In reality, users would not have to look down at a question several times during a conversation, so the fact that they sometimes had to do it under this experiment can be a source of error making the eye gaze steering conditions harder.

Some questions seemed a little hard for participants to answer, which was often due to the answer being hidden as an interposed sentence or the talker mumbling, which would also be real situation situations. However, if several hard questions were presented right after each other or grouped together, it might have been a source of error, since a condition could seem even harder than it was supposed to be. To avoid specific hard questions being presented in the same order or always in the same condition, everything in this experiment was balanced as much as possible, as described in section 4.5. Since there was not an equal amount of questions from each video or of each type, the questions were not balanced perfectly and video 10, for instance, was therefore shown more times than the other videos. This might not have had any influence since all questions were used an equal amount of time between all conditions. However, if a participant got a grouping of harder questions during soft eye gaze steering and a grouping of easy questions during hard eye gaze steering, hard eye gaze steering might have been easier due to the questions and not the eye gaze steering condition. For future experiments, it would be relevant to have the question tested for perceived difficulties, so this source of error could be eliminated.

6.3.3 Background noise

As described, it would be better to adjust the background noise to each participant, so the perceived difficulty could be equal amongst participants. This would probably have made participant 14 answer more questions correct since this participant had difficulties hearing anything in background noise due to his high SRT. In the opposite direction is participant 10, who could answer almost everything correct, because of a low SRT. Luckily, most participants could perform as expected with a fixed SNR at -3dB. An adjusted SNR for each participant would have been preferable, even though it was not possible for this master thesis. For future experiments, it would be interesting to develop material that could test hearing in noise with visual clues included.

Other sources of errors due to background noise could be that the noise was coming only from the front and that it was not reduced or enhanced in the same way as the speech signals. The background noise was decided only to come from the front because participants were using their own hearing aids, which could be in a directional setting. If this was the case for only some participants, they would not experience the background noise equal amongst participants if loudspeakers with background noise were placed behind them. If the background noise was enhanced and reduced according to participants' eye gaze, there would not have been a difference in SNR in each direction, since the background noise came from the same three directions as the speech signals. Even though the background noise may not have been acting very ecological, it still worked sufficiently to show that hard and soft eye gaze steering can help hear in a noisy environment.

6.3.4 Hardware and software

During the experiment, the biggest source of error was definitely the Dikablis eye tracker used in the experiment. It was very uncomfortable to wear on the nose, which a lot of participants complained about. It might have made participants rate the workload higher, especially the frustration-aspects, since they could have gotten even more frustrated due to the pain inflicted by the eye tracker. Furthermore, being physically annoyed by something on the nose can also have an impact on the general performance during the experiment.

Besides being uncomfortable, the eye tracker also needed to be calibrated several times during the experiment due to the eye gaze drifting and becoming more imprecise with time. Usually, this occurred because participants touched their noise due to the pain or because the eye tracker slowly slid down their nose during the experiment. This error did not have a big impact during the experiment since the gain curves were wide enough to understand that participants were looking at the talker, even though the eye gaze was a few centimeters off. However, this error resulted in eye gaze data collected being off to the sides, which was showed in section 5.4. On these histograms, it looks like the participants are looking in between talkers in the dialogue or beside the talkers a lot, which is probably due to the eye tracker measuring the gaze as being a little off to the side for some trials.

A source of errors could also be the program, which halfway through the experiment included an offset pushing the enhancement approximately 10 degrees to the right. Since the offset was corrected, the only thing it seemed to have disturbed was the eye gaze data, which had to be pushed 10 degrees to be understandable. This means that the histograms are even more imprecise, since there are both errors from the eye tracker being a little off, and the data collected after participant 8 that might not be corrected perfectly. Besides the histograms being influenced, it is not known if there is anything else the offset might have disturbed.

6.3.5 Order effect

The results indicate that participants did not get exhausted or close to giving up during the experiment. Furthermore, there is not seen any difference in the results related to when the no, hard, and soft eye gaze steering condition was presented. However, the familiarization is always seen to be easier, both in answering questions and when measuring workload. This might, of course, be because it is an easier condition without background noise, but it might also be because participants always get this condition as the first one. Nine participants commented that it only got harder after the first condition, and maybe this effect also comes from them being more awake in the first condition. Furthermore, before the first condition, they have not experienced pain from the eye tracker, and if this equipment hurt them throughout the experiment, it might have been more comfortable during the first condition than the next three.

6.3.6 Test participants

During the experiment, people with hearing loss and hearing aids were used as participants. This means that the real user group is tested and, therefore, the results are more trustworthy than if normal hearing participants were used. However, since the participants are connected to Eriksholm Research Centre they are often used as participants in experiments. They can therefore be used to participating in experiments, and moreover, they might be able to withstand more during two hours of testing than they would accept in the real world.

6.4 Eye gaze steering in hearing aids

As seen in the results from the exit-interview, participants like the idea of eye gaze steering in their hearing aids, and it would probably make them even happier about their hearing aids. It is furthermore seen in results from the speech intelligibility questions that participants perform better with help from eye gaze steering than no steering. From the results gathered from the eye gaze data is it seen that people actually look in the direction of the talkers and less on the monologue, which in this case presents information that participants are not interested in. Since participants look in the direction they want to listen, perform better with eye gaze steering, and like this form of steering, eye gaze steering is understood as a good way to steer a hearing aid. This is also supported by participants commenting that the monologue got less annoying during the eye gaze steering conditions, which was a win for them.

However, if eye gaze steering should work, it is important that eye gaze is measured correctly. In this experiment, participant 8 experienced eye gaze steering, which enhanced the talkers wrong. The participant then commented in the exit-interview that it was easier to do the task when he could hear all three talkers at the same time than during the eye gaze steering where the talkers were enhanced wrong. This, of course, makes very much sense, but it also underlines the fact that studies with a functioning prototype are essential, to see if eye gaze can be captured well enough for the concept of eye gaze steering to work in reality.

From the exit-interview it is also clear that participants generally prefer hard eye gaze steering over soft eye gaze steering, even though they did not all notice the enhancement during the experiment. This means that hypothesis 3 is also rejected. They prefer hard eye gaze steering because this gives them more control over the situation and their hearing loss. The two participants who prefer soft eye gaze steering say it is because with soft eye gaze steering their eyes do not have to jump around so much. However, since hard eye gaze steering enhances as much as possible in the direction the user is looking, while soft eye gaze steering divides the enhancement among the directions that are looked at over a period of time, hard eye gaze steering seems to generally help the hearing loss the most.

Even though hard eye gaze steering is generally preferred, this form of eye gaze steering only works really good when the user knows where the sound is coming from. This is generally the case in situations where the user can talk one-on-one, but not always in a multi-talker scenario, where several people are talking together. Furthermore, when using hard eye gaze steering the user can not look away from the target at any point, if the whole signal should be heard, which will probably not very user-friendly in the long run. In a multi-talker scenario, where the user is looking around at several people and listening to them, the soft eye gaze steering might still work very well. The same is true in a one-on-one conversation where the user probably looks at the talker, but then have the freedom to look away momentarily. This is due to soft eye gaze steering making it possible to get a smooth transition between the speech signals and not completely reduce the speech signal with momentarily gazes away from the targets, depending on how the parameters are defined.

6.4.1 Parameters of the eye gaze steering algorithms

In this experiment, the time measuring eye gaze during soft eye gaze steering was set to five seconds. As described earlier, this might have been either too much or not enough time. During the experiment, participants experienced the soft eye gaze steering enhancing too slow, when they shifted their gaze to another talker because they had focused five seconds or more at the first talker. Had the measuring time been about 10 seconds, then the other talker might still have been enhanced from the last time the participant looked in that direction, and therefore the conversation could have been easier and more relaxing to follow. But then again, if participants or users in general look more than ten seconds at something, then it will always take ten seconds for the next thing they look at to be enhanced fully, which can be way too much time and quite frustrating for users to wait for.

From the experiment it seemed clear that eye gaze steering in hearing aids could benefit the hearing aid user by actually reducing unwanted noise and compensating for the hearing loss in the direction attention is directed. However, when interpreting the results, it also seems like a mixture of hard and soft eye gaze steering could be preferable. For some participants the hard eye gaze steering was too hard, meaning that their eyes had to jump too much around and that the sudden enhancement and reduction of speech signals were confusing. Furthermore, the soft eye gaze steering seemed to be too slow for participants to feel in control. A combination of those two forms of eye gaze steering could be interesting to investigate. This could either be the before mentioned time measuring eye gaze that could be very short, for instance 1-2 seconds, where the eye gaze steering would work very much like hard eye gaze steering, but with softer shifts. Another possibility is to make a hard eye gaze steering beamformer with gain curves wide enough to include two talkers, which would probably also feel a little softer when shifting the eye gaze back and forth between two talkers.

Either way, more studies in this area would be preferable to understand when the soft eye gaze steering is working, or how the hard eye gaze steering could be optimized.

6.4.2 Answering the research question

To answer the research question stated in chapter 3 eye gaze steering of a beamformer actually helps the user by reducing unwanted noise or enhancing the sound in the direction they actually want to listen. From this experiment, hard eye gaze steering seems to be the better choice, but soft eye gaze steering still has a few liked features that users would also benefit from. To make the experience as user-friendly as possible, it could therefore be beneficial to investigate eye gaze steering as a combination of hard and soft eye gaze steering.

Chapter 7

Conclusion

Throughout the master thesis project, a lot of aspects related to eye gaze steering of a beamformer have been investigated, which will be concluded upon in this chapter. The two forms of eye gaze steering that have been investigated are hard eye gaze steering, which enhances the sound in the direction of the eye gaze momentarily, and soft eye gaze steering, which measures the eye gaze over time and enhances the sound in the directions that are being looked upon according to how much the individual direction is being looked at.

First of all, it can be concluded that eye gaze steering of a beamformer is better than no steering when measuring speech intelligibility, in the form of how many questions participants can answer correctly. This means that hypothesis 1(a) is accepted. However, hypothesis 2(a) is rejected since there is no significant difference between the two forms of eye gaze steering when measuring speech intelligibility. Furthermore, hard eye gaze steering seems to get more correct answers than soft eye gaze steering, which is pointing in the opposite direction of the hypothesis.

From the experiment it can also be concluded that no, hard, and soft eye gaze steering is scoring high when measuring workload through as NASA Task Load Index (TLX). The only difference is seen between familiarization and the other three conditions, where familiarization is scoring significantly lower in workload. This effect is also seen for mental demand when looking at each aspect of the NASA-TLX individually. With more experience and longer trial periods of the concept, maybe hard and soft eye gaze steering will be easier for users to manage, which will lower the workload. However, for this experiment, it is concluded that there is no difference between no, hard, and soft eye gaze steering when measuring workload, which is why hypotheses 1(b) and 2(b) are rejected.

When asked about it, most participants prefer the hard eye gaze steering, because they feel more in control with this form of eye gaze steering. Hypothesis 3 is therefore rejected. Even though hard eye gaze steering is generally preferred, this form of eye gaze steering only works really good when the user knows where the sound is coming from and if the user never looks away from the target. Since hard eye gaze steering is limited if the user does not know where the next speech signal is coming from, users might miss the first words of a sentence if the eye gaze is not shifted fast enough. Soft eye gaze steering was assumed to help with this problem, but in this experiment, no such effect was detected. It is concluded that the five seconds measuring eye gaze under soft eye gaze steering is either too long or too short to give any advantages or a natural feel.

From the experiment it is furthermore concluded that eye gaze steering in hearing aids could benefit the hearing aid user by actually reducing unwanted noise and compensating for the hearing loss in the direction attention is directed. Furthermore, the hard eye gaze steering of a beamformer seems to be the better choice. However, soft eye gaze steering still has a few features that participants would also benefit from. When interpreting the results, it also seems like a mixture of hard and soft eye gaze steering could be preferable. This could, for instance, be the time measuring eye gaze in soft eye gaze steering being very short, for instance 1-2 seconds, where the soft eye gaze steering would work much more like hard eye gaze steering with softer shifts. Or it is a possibility to make hard eye gaze steering with wider gain curves, which could include two talkers at a time, which would probably feel a little softer when shifting the eye gaze back and forth between talkers.

Even though the experiment worked great, improvements can still be made, and the user experience of eye gaze steering would be even better to measure if the experiment was done under more realistic settings.

Chapter 8

Future work

As described throughout the discussion in section 6, it would be beneficial to do more research in the area of eye gaze steering of beamformers for hearing aids. This is specifically regarding the parameters of both hard and soft eye gaze steering. For instance, it could be interesting to investigate how these two forms of eye gaze steering work if gain curves were wider, and therefore capturing more of the surroundings than seen in this experiment. Furthermore, it could be interesting to investigate how soft eye gaze steering would work when measuring eye gaze for a shorter or longer period of time. When it is known which parameters would make both hard and soft eye gaze steering work best, it would be beneficial to test these two against each other again, maybe in the same kind of setup or in a more ecological setup.

It would be beneficial to test the eye gaze steering of beamformers in more ecological surroundings, for instance outside of a laboratory or with present talkers when a working prototype is made for participants to try out. Lastly, it would also be beneficial to have participants try out the eye gaze steering for a longer period of time, so they would actually understand how this new technology could help them in their everyday life.

Chapter 9

Acknowledgement

I would like to thank both my university supervisor, Lars Bo Larsen, and my two company supervisors, Poul Hoang and Martha Shiell, for the support and guidance during my master thesis.

Especially Poul Hoang, Martha Shiell, and Sergi Rotger Griful should receive a great thanks for helping with the program and the experimental setup.

A big thanks should also go out to the 15 participants used in the experiment. Without you, none of this could have happened.

Finally, I would like to thank the employees at Oticon A/S and Eriksholm Research Centre for having me and contributing to a lovely work atmosphere.

Bibliography

American Speech-Language-Hearing Association. *Degree of Hearing Loss*. Available at: <https://www.asha.org/public/hearing/degree-of-hearing-loss/>.

Bernstein, J. G. W. and Grant, K. W., 2009. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearingand hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 05, 5, pp. 3358–3372.

Best, V., Roverud, E., Streeter, T., Mason, C. R., and Grald Kidd, J., 2017. The Benefit of a Visually Guided Beamformer in a Dynamic Speech Task. *SAGE*, 21, pp. 1–11.

Betkowski, B., 08 2018. *Why most people with hearing loss don't use hearing aids*. Available at: <https://medicalxpress.com/news/2018-08-people-loss-dont-aids.html>.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Acoustical Society of America*, 11, 110(5), pp. 2527–2538.

Buchan, J. N., Paré, M., and Munhall, K. G., 2008. The Effect of Varying Talkier Identity and Listening Conditions on Gaze Behavior During Audiovisual Speech Perception. *Brain Res*, 11, pp. 162–171.

Chakrabarty, S., Theirgart, O., and Habets, E. A. P., 2015. A baysian approach to spatial filtering and diffuse power estimation for joint dereverberation and noise reduction. *ICASSP*, pp. 753–757.

Cherry, E. C., 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *The journal of the acoustical society of America*, 09, 25(5), pp. 975–979.

Dillon, H. *Hearing Aids*. Thieme, 2 edition, 2012.

Elberling, C. and Worsøe, K. *Naar lyden bliver svagere - om hoerelse og hoereapparater*. Videnscenter for hørehandicap, 2 edition, 2007.

Ergoneers. *Dikablis Professional Wireless Eye-Tracking*. Available at: <https://www.ergoneers.com/en/landingpages/dikablis-professional-wireless-eye-tracking-system/>.

Field, A. and Hole, G. *How to Design and Report Experiments*. SAGE, 2003.

Field, A., Miles, J., and Field, Z. *Discovering Statistics Using R*. SAGE, 2012.

Harrison, M., 01 2018. Evaluating the Use of Steering a Hearing Aid in A Dynamic Multi-Talker Environment Using Body Signals. Speciale, University of Glasgow.

Hart, J., Onceanu, D., Sohn, C., Wightman, D., and Vertegaal, R. The Attentive Hearind Aid: Eye Selection of Auditory Sources for Hearing Imparied Users. In *Human-Computer Interaction – INTERACT 2009*, pp. 19–35, 2009.

Heldner, M. and Edlund, J., 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38, pp. 555–568.

Hoang, P., Jensen, J., de Haan, J., Tan, Z.-H., and Lunner, T. Robust Bayesian and Maximum a Posteriori Beamforming for Hearing Assistive Devices. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE Global Conference on Signal and Information Processing (GlobalSIP). Proceedings. IEEE Signal Processing Society, 9 2019.

Human Performance Research Group and NASA Ames Research Center. *NASA TASK LOAD INDEX (TLX)*.

Kidd, G., Favrot, S., Desloge, J. G., Streeter, T. M., and Mason, C. R., 2013. Design and preliminary testing of a visually guided hearing aid. *Acoustical Society of America*, 03, 133(3).

Laerd Statistics, 2018. *One-way repeated measures MANOVA in SPSS Statistics*. Available at: <https://statistics.laerd.com/spss-tutorials/one-way-repeated-measures-manova-using-spss-statistics.php>.

Launer, S., Zakis, J. A., and Moore, B. C. J. *Hearing Aids: Hearing Aid Signal Processing*, volume 56. Springer, 2016.

McLeod, S. A., 06 2019. *What does effect size tell you?* [Simply Psychology]. Available at: <https://www.simplypsychology.org/effect-size.html>.

Moore, B. C. J. *An Introduction to the Psychology of Hearing.* Brill, 6 edition, 2013.

Nilsson, M., Soli, S. D., and Sullivan, J. A., 1994. Development of the Hearing In Noise Test for the measurement of speech reception threshold in quiet and in noise. *The Journal of the Acoustical Society of America*, 03.

Phycho Hawks, 10 2010. *Effect size for Analysis of Variance (ANOVA).* Available at: <https://psychohawks.wordpress.com/2010/10/31/effect-size-for-analysis-of-variables-anova/>.

Vatikiotis-Bateson, E., Eigsti, I.-M., and Yano, S., 1994. Listener eye movement behavior during audioivisual speech perception. *International Conference on Spoken Language Processing*, pp. 527–530.

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., and Munhall, K. G., 1998. Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, 60 (6), pp. 926–940.

Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A., 2001. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Then Meets the Eyes. *CHI*, 3(1), pp. 301–308.

Vicon. *We go beyond motion.* Available at: <https://www.vicon.com/>.

World Health Organization, 03 2019. *Deafness and hearing loss.* Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.

Xia, J., Kalluri, S., Micheyl, C., and Hafter, E., 2017. Continued search for better prediction of aided speech understanding in multi-talker environments. *Acoustical Society of America*, 142(4), pp. 2386–2399.

Appendices

Appendix A

Overview over Electronic Appendix

A.1 Example of audiovisual material

An example of a video clip from the audiovisual material can be found in (ElectronicAppendix/ExampleVideoClip.mp4). This example is from video 10.

A.2 Balance

The balance of questions and conditions for each participant can be found in (ElectronicAppendix/Balance).

A.3 Notes and answers during the experiment

The notes and answers to both questions and exit-interview written during the experiment for each participant can be found in (ElectronicAppendix/Results/AnswerSheet.xlsx)

A.4 Workload scores and speech intelligibility in percent

The workload scores with weighting and the percentage of correct answers can be found in (ElectronicAppendix/Results/Workload_SI.xlsx)

A.5 Eye gaze data

The eye gaze data collected during each trial for each participant can be found in (ElectronicAppendix/Results/EyeGazeData). Data concerning only the eye gaze data is called gazeRelAng in the .mat file.

Appendix B

Questions for measuring speech intelligibility

In the following chapter, all questions used in the experiment are presented. Questions with two stars (**) have the answer or part of the answer given as the first thing in the next talker's sentence. Questions with three stars (***) is a "backward" question, where the answer comes before the words that are presented in the questions (Q: What color is the cat? A: There is a brown cat). Questions without any stars are normal questions with the question first, and the answer afterward.

Video 5

Q1: Hvilken farve hår har den omtalte kvinde? Brunt **
Q3: Hvad ligger ketcherne under bordet i? En kasse **
Q5: Hvad ligner boksen? Et busstoppested
Q7: Hvilken farve dæk synes hun taxaen har? Sorte **
Q8: Hvilken farve er sæderne i taxaen? Orange ***
Q9: Hvad har den omtalte kvinde med? En pink badering **
Q10: Hvad sidder på forsædet i jeepen? En hund ***
Q11: Hvor mange mennesker skubber til bilen? To ***
Q12: Hvad siger manden på hendes tegning? Ik' noget
Q14: Hvilken farve er spanden? Rød **

Video 9

Q15: Hvilken by synes hun det ligner på kortet? Odense
Q17: Hvilken farve er slagterbutikken? Blå
Q19: Hvilken farve er den forvoksede parfume flaske? Babyblå **
Q20: Hvad står med rødt på det hvide skilt? Apotek
Q22: Hvad er der ovenover apoteket, udover en sky? En sol **
Q23: Hvad kalder hun den omtalte kvinde i billedet? En travl buisness woman**
Q25: Hvilken farve er hunden? Brun
Q26: Hvilken farve er bilen? Rød ***

Q28: Hvilken farve er det lille logo? Gult **
Q32: Hvilken farve er styret? Sort **
Q33: Hvad har de to mænd fået på lagkagehuset? Romkugler of caffe latte
Q34: Hvor sidder håndtagene? Meget langt nede
Q35: Hvilken farve er døren? Karry **

Video 10

Q36: Hvor mange søm har det hvide skilt? To ***
Q37: Hvilken slags hat har manden på hovedet? En grøn hat/En baret **
Q39: Hvor mange hvide aviser har han på tegningen? Seks ***
Q40: Hvilken farve er fuglen? Gul
Q41: Hvad vender hunden numsen mod? Skiltet
Q43: Hvor er kødtærten fra? Bageren
Q44: Hvad er pigens yndingsfarve? Pink/rød
Q45: Hvilken farve kjole er der på billedet? Blå ***
Q48: Hvilke farver er det kirkelige vindue? Gult og orange
Q49: Hvilken farve ramme har den grå dør? Grøn ***
Q50: Hvilken farve synes han jakkesættet er? Laksefarvet
Q51: Hvor mange par sko er der på hans tegning? Tre par **
Q54: Hvilket humør er damen, som lige er blevet mor, i? Rasende **
Q56: Hvilken butik er der bag den omtalte dreng? Helsekost butik

Video 11

Q59: Hvilken farve er taget på butikken? Grøn ***
Q60: Hvilken farve er manden på hendes kort? Hvid **
Q61: Hvad mener han fuglene hakker i? Korn
Q64: Hvilken dør er åben ind til apoteket? Den venstre dør **
Q65: Hvad sælger den gullige butik? Legetøj
Q66: Hvad har den højre bamse på, på hendes billede? En vest
Q68: Hvad gør katten? Sover
Q69: Hvilken farve har bænken? Mosgrøn **

Appendix C

Translation of NASA-TLX questions

Six English and Danish speaking persons were asked to translate the questions from the NASA-TLX questionnaire, which are as following:

- How mentally demanding was the task?
- How physically demanding was the task?
- How hurried or rushed was the pace of the task?
- How successful were you in accomplishing what you were asked to do?
- How hard did you have to work to accomplish your level of performance?
- How insecure, discouraged, irritated, stressed, and annoyed were you?

What the six people answered were:

Mental demanding:

1: Hvor mentalt krævende var opgaven?
2: Hvor mentalt krævende var opgaven?
3: Hvor mentalt krævende var opgaven?
4: Hvor mentalt udfordrende var opgaven?
5: Hvor mentalt krævende var opgaven?
6: Hvor mentalt krævende var opgaven?

Physical demanding:

1: Hvor fysisk krævende var opgaven?
2: Hvor fysisk krævende var opgaven?
3: Hvor fysisk krævende var opgaven?
4: Hvor fysisk udfordrende var opgaven?
5: Hvor fysisk krævende var opgaven?
6: Hvor fysisk krævende var opgaven?

Temporal demand:

1: Hvor forhastet eller fremskyndet var tempoet af opgaven?
2: Hvor hastet eller presseret var opgavens tempo?
3: Hvor stressende/skyndt var tempoet på opgaven?
4: Hvor forhastet eller forjaget var tempoet på opgaven?
5: Hvor forhastet var tempoet på opgaven?
6: Hvor hurtigt var tempoet af opgaven?

Performance:

1: Hvor vellykket var du til at udføre det du blev bedt om at gøre?
2: Hvor succesfuld var du i at opnå det, du blev bedt om?
3: Hvor succesfuld var du til at løse hvad du blev bedt om at gøre?
4: Havde du succes med at udføre de ting du blev spurgt om at gøre?
5: Hvor succesfuld var du i at opnå det du blev bedt om at gøre?
6: Hvor succesfuld var du i at gennemføre opgaven?

Effort:

1: Hvor hårdt blev du nødt til at arbejde for at opnå dit præstationsniveau?
2: Hvor hårdt var du nødt til at arbejde for at opnå dit præstationsniveau?
3: Hvor hårdt havde du brug for at arbejde for at løse opgaven og opnå dit præstationsniveau?
4: Hvor mange kræfter skulle du lægge i, for at opnå dit performance niveau?
5: Hvor hårdt skulle du arbejde for at opnå dit præstationsniveau?
6: Hvor hårdt skulle du arbejde for at gennemføre dit niveau af udøvelse?

Frustration:

1: Hvor usikker, umotiveret, irriteret, stresset og irritabel var du?
2: Hvor usikker, modløs, irriteret og stresset var du?
3: Hvor usikker, modløs, irriteret, stresset og irriteret var du?
4: Hvor usikker, modløs, irriteret, stresset og generet var du?
5: Hvor usikker, afskrækket, irriteret, stresset og generet var du?
6: Hvor usikker, afmodet, irriteret, stresset og irriteret var du?

Since the translations were so similar, the translation given to test participants in need was as the following:

- Hvor mentalt krævende var opgaven?
- Hvor fysisk krævende var opgaven?
- Hvor forhastet eller forjaget var tempoet på opgaven?
- Hvor succesfuld var du i at gennemføre opgaven?
- Hvor hårdt var du nødt til at arbejde for at opnå dit præstationsniveau?
- Hvor usikker, modløs, irriteret, stresset og generet var du?

Appendix D

The eye gaze steering algorithms

For this experiment, speech enhancement systems that conceptualize how a beamformer would work is designed. To understand the mathematics of the hard and soft eye gaze enhancement system, a scenario can be set up. Figure D.1 shows the scenario related to the experiment, where three talkers are shown on a curved TV screen and a hearing aid user is listening to them. Three loudspeakers behind the TV are playing the speech signal from each talker, and three smaller loudspeakers on the floor in front of the TV are playing background noise. The speech signal from the talkers is defined as s1, s2, and s3, while the background noise is defined as v1, v2, and v3. The noise can also be noted as:

$$V(n) = v_1(n) + v_2(n) + v_3(n) \tag{D.1}$$

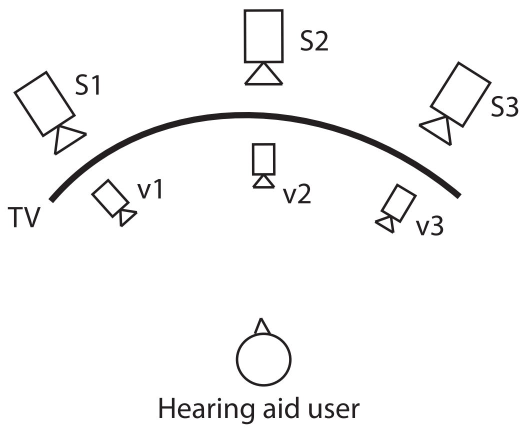


Figure D.1: The figure shows a setup with three talkers on a curved TV screen, a hearing aid user, loudspeakers to enhance the talkers' speech signals and loudspeakers below the TV in the front playing background noise.

For simplicity, the hearing aid user is understood as a single microphone placed in the middle of the user's head. If the sound reaching the hearing aid user is measured in discrete time, it can be modeled as:

$$x(n) = s_1(n) + s_2(n) + s_3(n) + V(n) = \sum_{i=1}^{3} s_i(n) + V(n) \tag{D.2}$$

Where *x(n)* is the unprocessed signal reaching the hearing aid user, $s_1(n)$, $s_2(n)$, and $s_3(n)$ are the signals reaching the hearing aid user from each talker, and *V(n)* is the background noise reaching the user.

D.0.1 Hard eye gaze steering

To enhance the talkers from figure D.1 with hard eye gaze steering, gain curves simulating a beamformer for each talker are needed. An example of gain curves can be seen in figure D.2. Θ is the span of angles within -180 degrees and 180 degrees, wherein eye gaze can lie.

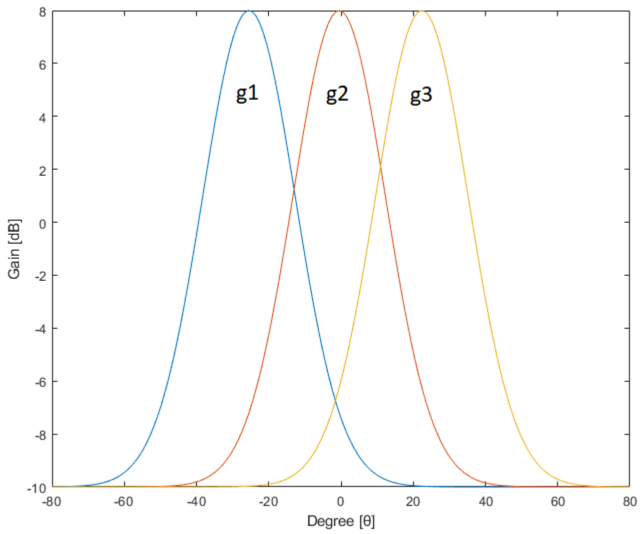


Figure D.2: The figure shows an example of gain curves for each talker, if looked in their direction.

If gain curves are determined, and eye gaze, $\phi(n)$, is measured, the enhanced sound is found by using the following algorithm:

$$y(n, \phi(n)) = g_1(\phi(n)) \cdot s_1(n) + g_2(\phi(n)) \cdot s_2(n) + g_3(\phi(n)) \cdot s_3(n) + V(n) \tag{D.3}$$

or

$$y(n, \phi(n)) = \sum_{i=1}^{3} g_i(\phi(n)) \cdot s_i(n) + V(n) \tag{D.4}$$

Here $y(n, \phi(n))$ is the processed signal reaching the hearing aid user, $\phi(n)$ is the measured eye gaze, and $g_i(\phi(n))$ is the gain for the specific eye gaze direction, which can be measured on the gain curves.

D.0.2 Soft eye gaze steering

To understand the mathematics behind the soft eye gaze beamformer, the scenario in figure D.1 can be used again. An assumption for the soft eye gaze steering is that the hearing aid user will be looking at the directions of the target talkers. The soft eye gaze steering works by calculating the enhancement for all direction and multiply it with the probability of the user will look in those directions. This can be modeled as:

$$y_B(n) = \sum_{\text{j=1}}^{J} P(\theta_j|\Phi) \cdot y(n, \theta_j) \tag{D.5}$$

Where $y_B(n)$ is the enhanced sound using the Bayesian approach, [Hoang et al., 2019], J is the number of directions, $y(n, \theta_j)$ is the beamformer for direction j, θ_j is the target direction, Φ is the eye gaze directions measured over a time period, and $P(\theta_j|\Phi)$ is the probability that the target talker is being looked upon when Φ is measured. $P(\theta_j|\Phi)$ can be depicted as a histogram, where the frequency of the eye gaze direction in each direction is depicted, see figure D.3. For the sum of probabilities to be equal to one, the histogram has to be normalized. $y_B(n)$ can also be described as:

$$y_B(n) = \sum_{\text{j=i}}^{J} P(\theta_j|\Phi) \cdot (\sum_{\text{i=1}}^{3} g_i(\theta_j) \cdot s_i(n) + V(n)) \tag{D.6}$$

or

$$y_B(n) = \sum_{\text{i=1}}^{3} s_i(n) \cdot \sum_{\text{j=1}}^{J} P(\theta_j|\Phi) g_i(\theta_j) + V(n) \tag{D.7}$$

Here $s_i(n)$ is the specific talker and *V(n)* is the background noise. In regards to indexing, i is the index of the i'th talker and j is related to the j'th direction. To simplify the equation,

$$h_i(\Phi) = \sum_{\text{j=1}}^{J} P(\theta_j|\Phi) g_i(\theta_j) \tag{D.8}$$

where $h_i(\Phi)$ is the final gain used to enhance the i'th talker speech signal. $h_i(\Phi)$ is the summarized histogram you get, when the probability, or histogram, of eye gaze over time is multiplied with the gain curves for the enhancement. From that, a simplified model of the processed soft steering signal reaching the hearing aid user can therefore be modeled as:

$$y_B(n) = \sum_{\text{i=1}}^{3} s_i(n) \cdot h_i(\Phi) + V(n) \tag{D.9}$$

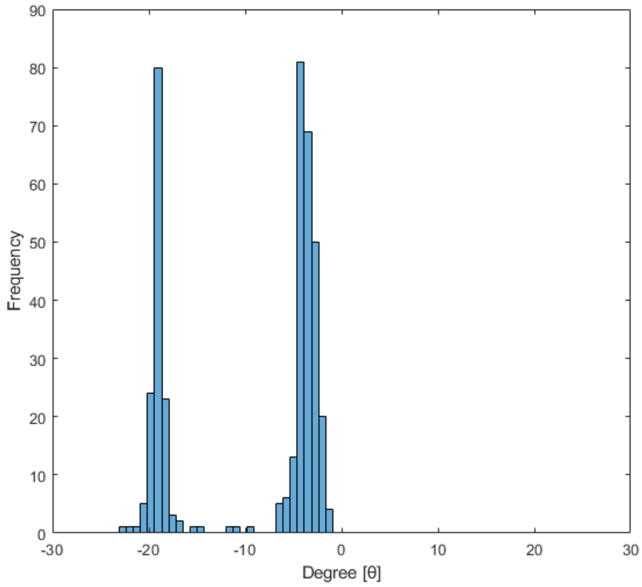


Figure D.3: The figure shows an example of a histogram over eye gaze collected over time. Note that the histogram is not normalized.

Appendix E

Audiograms

Figure E.1 and E.2 shows the average audiogram from the participants as the thick blue line, whereas the black dotted line is the audiogram from each participant. Figure E.1 shows the audiogram of the right ear, and figure E.2 shows the audiogram of the left ear. The participants hearing loss streched from slight to moderately severe, [American Speech-Language-Hearing Association].

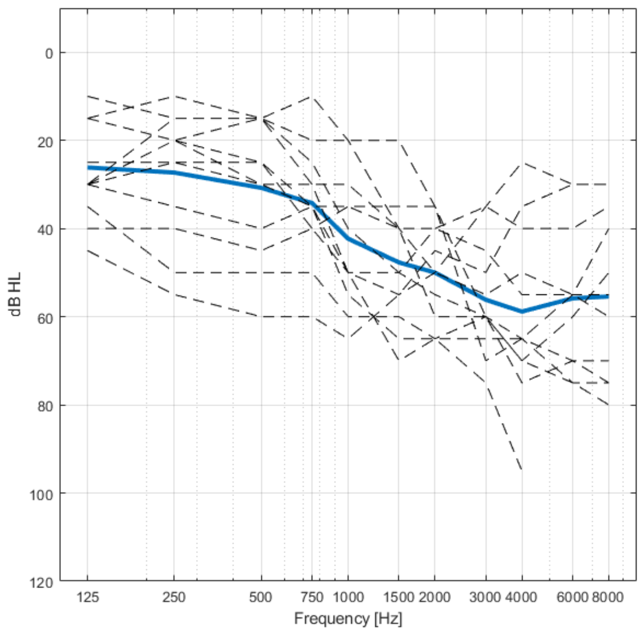


Figure E.1: Audiogram from the right ear of participants. The black dotted lines show the audiograms from each participant, while the thick blue line shows the average audiogram of participants' right ear. On the x-axis the eight difference frequencies measures are shown and on the y-axis the dB hearing level (HL) is shown.

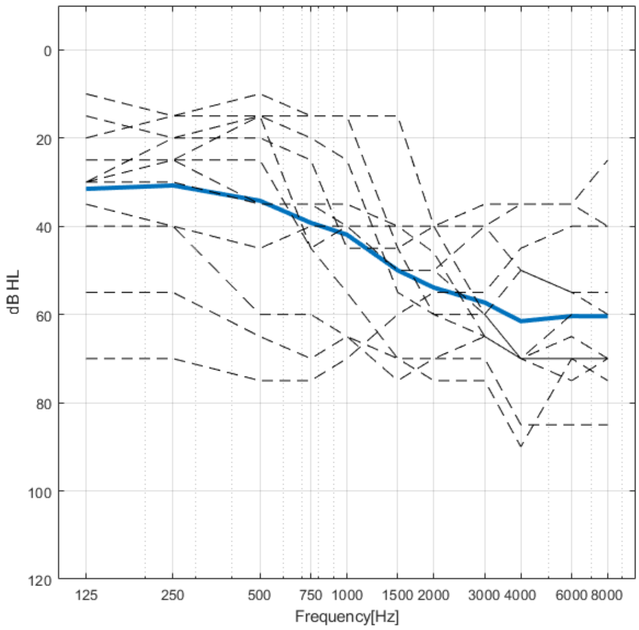


Figure E.2: Audiogram from the left ear of participants. The black dotted lines show the audiograms from each participant, while the thick blue line shows the average audiogram of participants' left ear. On the x-axis the eight difference frequencies measures are shown and on the y-axis the dB hearing level (HL) is shown.

Appendix F

Manuscript for experiment

- Hi and welcome!
- For how long have you been using a hearing aid?
- At which setting is your hearing aid? Okay, just keep it on that setting throughout the test.
- Today we are going to investigate how much it can benefit your user experience when the hearing aid enhances the sound in the direction you are looking.
- The procedure for today is as follows:

The procedure is explained from a printed overview of the procedure

- Now we will run a hearing in noise test (HINT). Maybe you have tried that before? If not, you will be presented for some sentences in noise. You just have to repeat the sentences as good as possible to me.

HINT programme is running

- Okay, now that we know how your speech intelligibility in noise is, we can calibrate the eye/head tracker and begin with the real testing.

Calibrating the eye/head tracker

- During the test you will be shown different video clips. On the videos, there will be 3 actors per clip – 4 different actors in total. One of the actors are telling different stories to the woman behind the camera, the other two are playing a game where they have to talk about what is on their individual pictures and try to find the things that are different from each other's pictures.
- The two actors playing the game are having a conversation, that you should follow.
- You will be handed a question on paper for every video clip, which is related to the two actors' conversation. You will be dropped in the middle of a conversation, and then you should try to follow the conversation.

- When you know the answer to the question handed to you, you should press the green button. Then you can tell me the answer afterward. The answer should be based only on their conversation in the clip. I mostly care about the accuracy of the answer, but I also want as fast a response as possible, so press the green button as soon as you know the answer.
- When the question mentions "him" or "her" it is referred to the people having the conversation. If the question mentions "the mentioned woman/man", then it is the woman/man in their conversation.
- If you do not know the answer, then just tell me that. You do not need to guess.
- When you have answered 11 questions, you will get a questionnaire about your workload to answer. If you need an explanation of the words, then just ask.
- It is important that you use your eye gaze as natural as possible.
- Do you have any questions? Otherwise, do not hesitate to ask during the test.

Familiarization session with questions and NASA-TLX

- Now we try another condition. Remember to use your eye gaze as natural as possible.

No/hard/soft eye gaze steering condition with questions and NASA-TLX

- Now we try another condition. Remember to use your eye gaze as natural as possible.

No/hard/soft eye gaze steering condition with questions and NASA-TLX

- Now we try another condition. Remember to use your eye gaze as natural as possible.

No/hard/soft eye gaze steering condition with questions and NASA-TLX

- Okay, great! That was all the conditions.
- Now I would like you to look at these cards one at a time, and circle the aspect you find most important of these two.

Weighing NASA-TLX aspects

- Before I let you go, I have a few last questions to ask you.

Exit-interview

- Thank you so much for your time today!

Appendix G

Assumptions for one-way repeated measures MANOVA analysis

In this appendix, some of the assumptions for the one-way repeated measures MANOVA and repeated measures ANOVA are tested. On table G.1 Shapiro-Wilk normality test and Mauchly's test for sphericity is shown. Data seems to generally be normally distributed and maintaining sphericity for all aspects, except the temporal demand aspect, where violations for both assumptions are violated. On figure G.1, G.2, G.3, and G.4 scatterplots are showing the relationship between the aspects for each condition. From the results, it seems there is no linear relationship between each pair of dependent variables for each condition, which means the assumption regarding linear relationship is violated.

	Familiarization	No steering	Hard steering	Soft steering
	Mental demand			
Shapiro-Wilk normality test	W = 0.96 p = 0.764	W = 0.90 p = 0.144	W = 0.93 p = 0.433	W = 0.93 p = 0.335
Mauchly's test for sphericity	W = 0.566 p = 0.357			
	Physical demand			
Shapiro-Wilk normality test	W = 0.89 p = 0.112	W = 0.88 p = 0.098	W = 0.89 p = 0.134	W = 0.94 p = 0.523
Mauchly's test for sphericity	W = 0.799 p = 0.825			
	Temporal demand			
Shapiro-Wilk normality test	W = 0.83 p = 0.023	W = 0.96 p = 0.844	W = 0.88 p = 0.086	W = 0.84 p = 0.028
Mauchly's test for sphericity	W = 0.260 p = 0.023			
	Performance			
Shapiro-Wilk normality test	W = 0.93 p = 0.360	W = 0.74 p = 0.002	W = 0.90 p = 0.140	W = 0.92 p = 0.287
Mauchly's test for sphericity	W = 0.628 p = 0.360			
	Effort			
Shapiro-Wilk normality test	W = 0.94 p = 0.490	W = 0.87 p = 0.065	W = 0.93 p = 0.358	W = 0.90 p = 0.167
Mauchly's test for sphericity	W = 0.568 p = 0.360			
	Frustration			
Shapiro-Wilk normality test	W = 0.96 p = 0.997	W = 0.94 p = 0.527	W = 0.97 p = 0.871	W = 0.91 p = 0.230
Mauchly's test for sphericity	W = 0.092 p = 0.000			

Table G.1: Table showing results from running a Shapiro-Wilk normality test and a Mauchly's test for sphericity. Assumptions are generally met for normality and sphericity, except under the temporal demand aspect, since the p-value shows a significant difference.

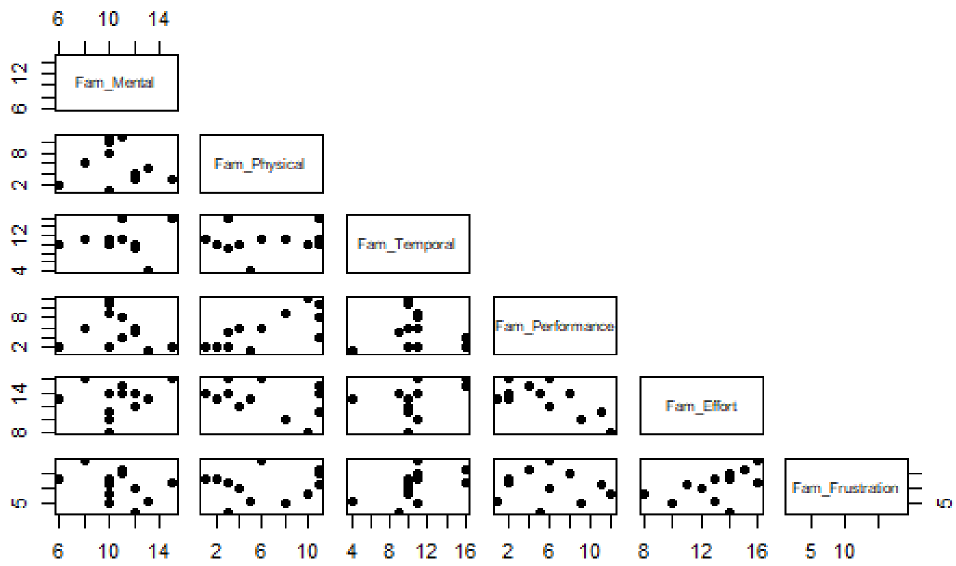


Figure G.1: Scatterplot matrix showing the relationship between the dependent variables, the aspects, for the familiarization condition.

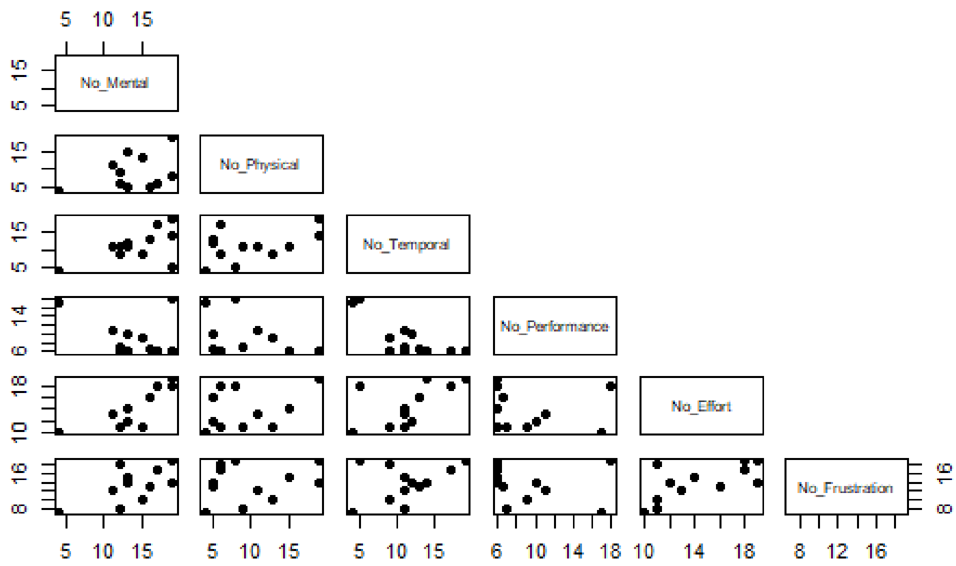


Figure G.2: Scatterplot matrix showing the relationship between the dependent variables, the aspects, for the no steering condition.

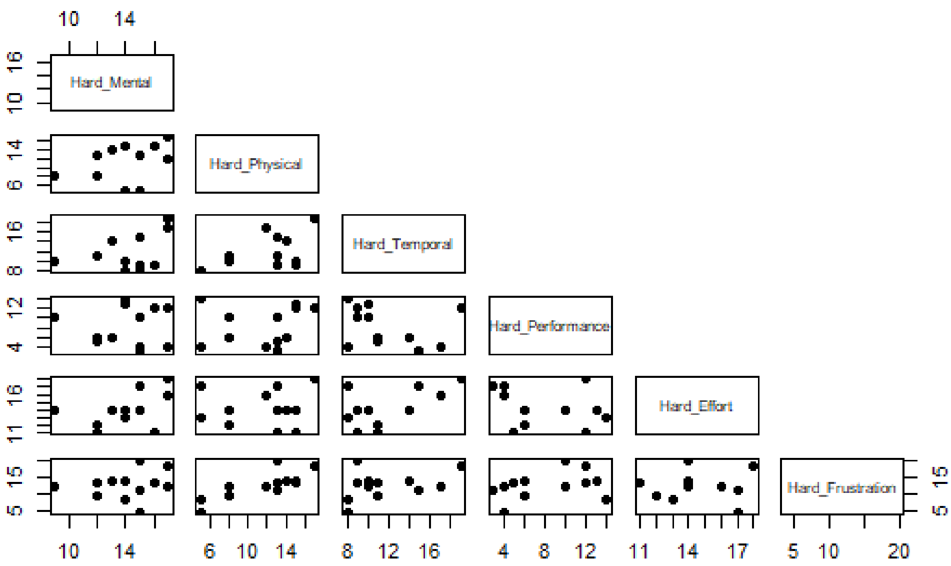


Figure G.3: Scatterplot matrix showing the relationship between the dependent variables, the aspects, for the hard eye gaze steering condition.

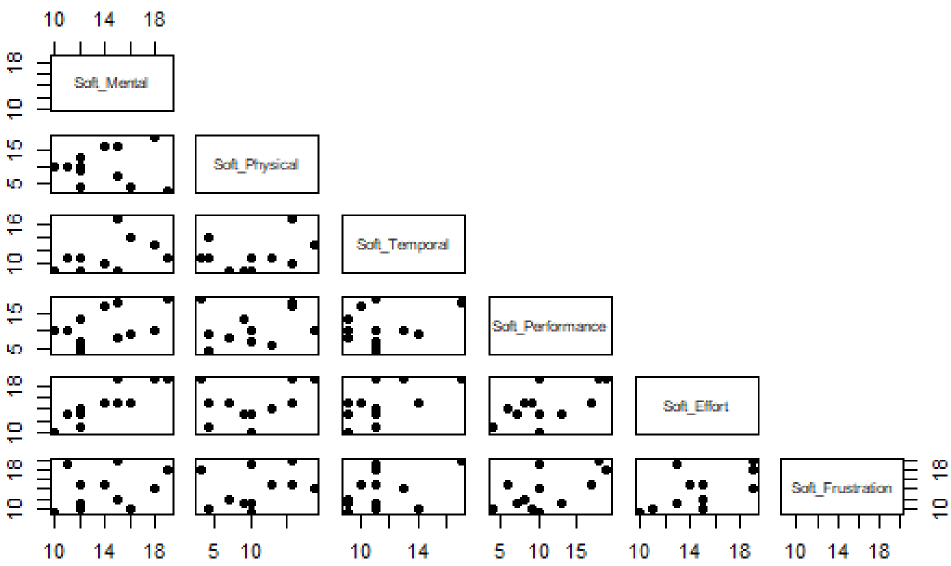


Figure G.4: Scatterplot matrix showing the relationship between the dependent variables, the aspects, for the soft eye gaze steering condition.

Appendix H

Eye gaze histograms

This appendix shows histograms over eye gaze data from each condition during video 5, 9, and 10. The eye gaze data from participant 9 to 15 are corrected with 10 degrees, which might not be precise enough for all data. The histograms should therefore not be used to anything but see in which directions participants generally looked.

H.1 Video 5

Familiarization

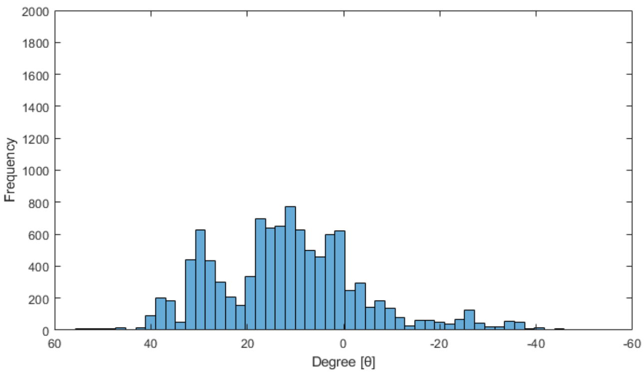


Figure H.1: The histogram shows eye gaze data from the participants during video 5 under the familiarization round. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

No steering

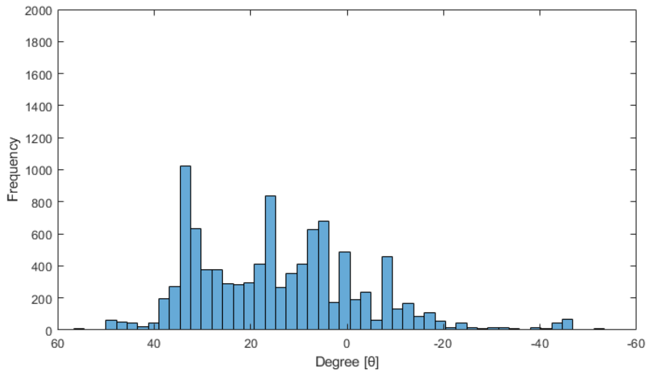


Figure H.2: The histogram shows eye gaze data from the participants during video 5 under the no steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

Hard eye gaze steering

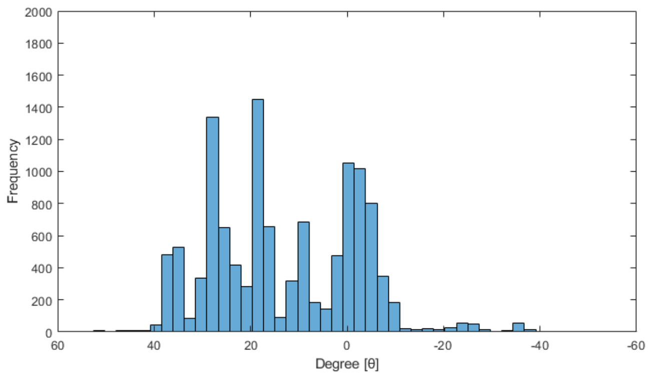


Figure H.3: The histogram shows eye gaze data from the participants during video 5 under the hard eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

Soft eye gaze steering

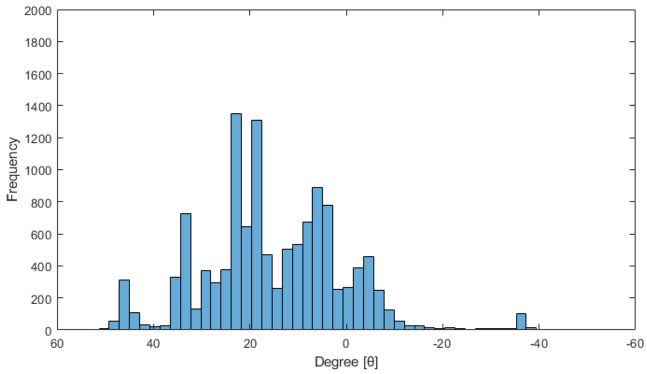


Figure H.4: The histogram shows eye gaze data from the participants during video 5 under the soft eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and middle talker.

H.2 Video 9

Familiarization

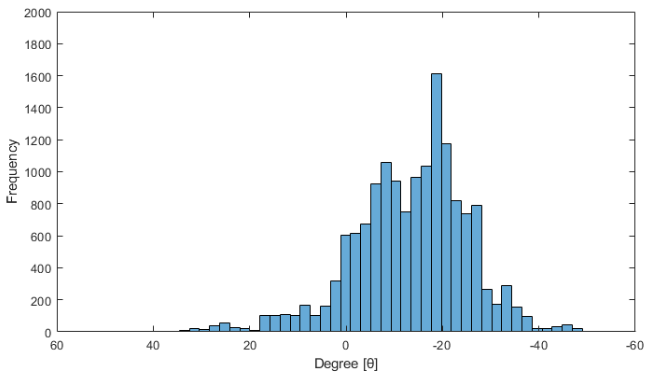


Figure H.5: The histogram shows eye gaze data from the participants during video 9 under the familiarization round. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the middle and right talker.

No steering

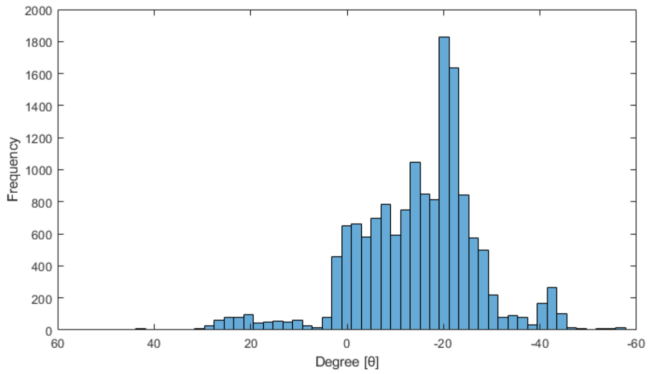


Figure H.6: The histogram shows eye gaze data from the participants during video 9 under the no steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the middle and right talker.

Hard eye gaze steering

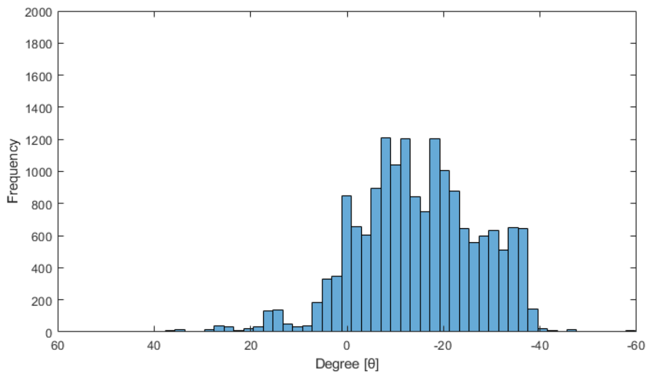


Figure H.7: The histogram shows eye gaze data from the participants during video 9 under the hard eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the middle and right talker.

Soft eye gaze steering

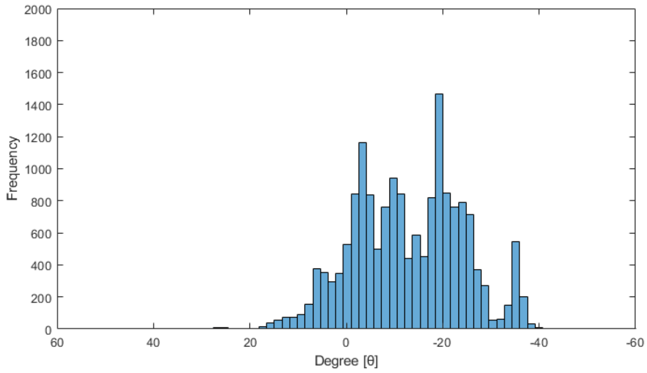


Figure H.8: The histogram shows eye gaze data from the participants during video 9 under the soft eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the middle and right talker.

H.3 Video 11

Familiarization

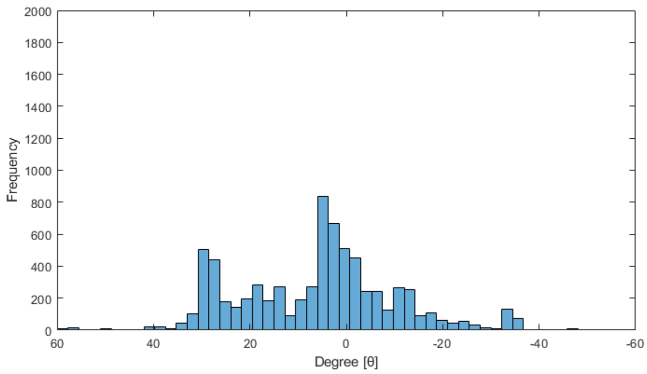


Figure H.9: The histogram shows eye gaze data from the participants during video 11 under the familiarization round. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and the middle talker.

No steering

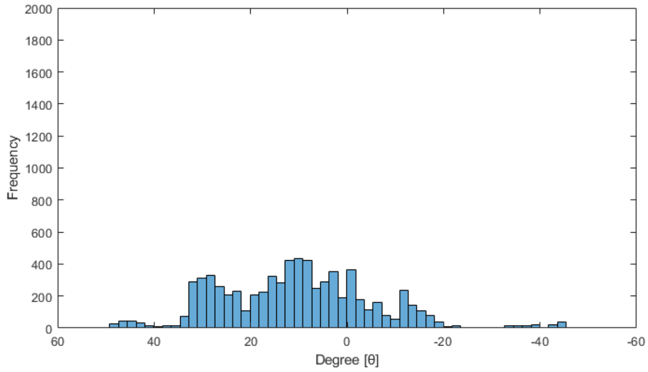


Figure H.10: The histogram shows eye gaze data from the participants during video 11 under the no steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and the middle talker.

Hard eye gaze steering

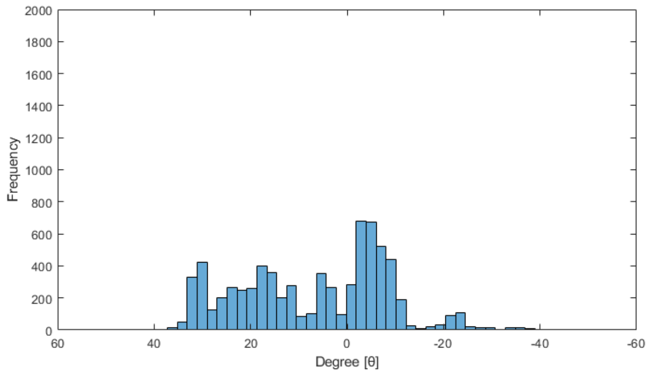


Figure H.11: The histogram shows eye gaze data from the participants during video 11 under the hard eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and the middle talker.

Soft eye gaze steering

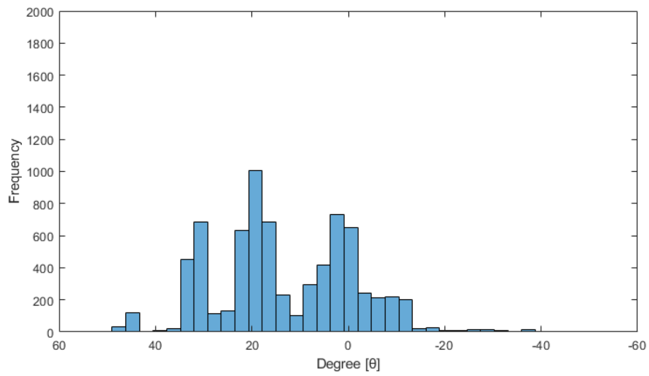


Figure H.12: The histogram shows eye gaze data from the participants during video 11 under the soft eye gaze steering condition. The left talker is placed approximately at 20 degrees, the middle talker approximately at 0 degrees, and the right talker approximately at -20 degrees. The dialogue is between the left and the middle talker.